IMPORTANCE OF DEFINING ROI:

A SEMI- AUTOMATED ALGORITHM FOR

PREDICTING IMPORTANCE MAPS

By

VAMSI KRISHNA KADIYALA

Bachelor of Engineering

Andhra University

Vizag, AP

2005

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2007

IMPORTANCE OF DEFINING ROI:

A SEMI- AUTOMATED ALGORITHM FOR

PREDICTING IMPORTANCE MAPS

Thesis Approved:

Dr Damon Chandler

Thesis Adviser

Dr Guoliang Fan

Dr Keith Teague

Dr. A. Gordon Emslie

Dean of the Graduate College

TABLE OF CONTENTS

Chapter **Page**

1 Introduction **1**

1.1 Motivation **2**

1.2 Previous work **3**

1.3 Outline **9**

2 Experiment-1 **10**

2.1 Methods **11**

2.1.1 Apparatus **11**

2.1.2 Stimuli **11**

2.1.3 Procedures **18**

2.1.4 Subjects **20**

2.1.5 Contrast Metric **21**

2.2 Results and Analysis **22**

Chapter **Page**

3 Experiment-2 **32**

3.1 Methods **32**

3.1.1 Apparatus **32**

3.1.2 Stimuli . . . . . **.33**

3.1.3 Procedures **38**

3.1.4 Subjects **40**

3.2 Results **40**

4 Experiment-3 **43**

4.1 Methods **43**

4.1.1 Apparatus **43**

4.1.2 Stimuli . . . . . **..44**

4.1.3 Procedures . . . . . **.. 47**

4.1.4 Subjects **47**

4.2 Results and Algorithm **47**

4.2.1 Algorithm **48**

4.3 Comparison **58**

5 Conclusions and Future Work **62**

BIBLIOGRAPHY

LIST OF FIGURES

Figure Page

1.1 (a) Image of a horse (b) Output of edge detector 4

1.2 (a) Input image (b) Output importance map 4

1.3 (a) Dog image (b) Output of Itti's algorithm 7

1.4 (a) Image of man and dog (b) Output of Osberger's algorithm . . . 8

2.1 512×512 grayscale images from the McGill database 12

2.2 Binary masks created by the author 14

2.3 Representative distorted images used in Experiment 1 17

2.4 Screenshot of experimental set-up 18

2.5 Subjective ratings of fidelity for images containing Gaussian white noise (High) 23

2.6 Subjective ratings of fidelity for images containing Gaussian white noise (Low) 24

2.7 Subjective ratings of fidelity for images containing wavelet subband quantization distortion (High) 25

2.8 Subjective ratings of fidelity for images containing wavelet subband quantization distortion (Low) 26

2.9 Subjective ratings of fidelity for images containing Gaussian blurring (High) 27

2.10 Subjective ratings of fidelity for images containing Gaussian blurring (High) 28

Figure Page

3.1 Background from van Hateren Database 33

3.2 (a) Dog (b) Soldier (c) Hydrant 34

3.3 Five images showing variation in size 35

3.4 Five images showing variation in Location 36

3.5 Five images showing variation in Blur 37

3.6 Five images showing variation in Contrast 38

3.7 Screenshot of experimental set-up 39

3.8 Subjective ratings of visual interest 41

4.1 Fifteen different images used in the experiment 45

4.2 Masks showing segmented regions for the images 46

4.3 Sigmoidal fit for Size factor 51

4.4 Sigmoidal fit for Blur factor 52

4.5 Sigmoidal fit for Contrast factor 54

4.6 Graph for the best fit 55

4.7 Graph for Osberger's Algorithm 56

4.8 Graphs with no Category 57

4.9 Output of proposed Algorithm 59

4.10 Output of Osberger's Algorithm 60

4.11 Output of Laurent Itti's algorithm 61

LIST OF TABLES

Table Page

4.1 Importance of each of the categories **49**

4.2 Parameters in equation three **50**

4.3 Parameters in equation four **52**

4.4 Parameters in equation five **53**

4.5 Optimal weights for the categories **55**

4.6 Optimal weights for Osberger's factors **56**

4.7 Optimal weights with no category **57**

CHAPTER 1

Introduction

The main objective of this research is to establish the importance of defining "regions of interest" in an image for improving the perceived fidelity of the image and propose an algorithm that would find these regions automatically in an image. The "regions of interest" or popularly known as the "ROI" can be defined as the objects or regions to which a viewers attention is naturally drawn. In simple terms it is the most important regions of the image. The term non-ROI can be very confusing. It usually refers to the regions in the image like the sky or the ground which do not attract our attention when we look at the image.

The problem of identifying regions of interest in a given image is an important one considering the wide range of applications it can be applied to. This concept can be affectively applied to improve the image quality assessments. The distortion metrics that are presently being used are not very affective in determining the different distortions present in an image accurately. This can be attributed to the fact that they do not take into account where the distortions occur in an image. Of late it has been accepted that distortions present in the ROI have a greater impact in determining the quality rating of an image when compared to distortions in non-ROI. Thus developing a new metric by taking into account where these distortions actually occur could really be useful.

The compression techniques that are presently being employed use uniform quantization for all the regions of an image. But the results obtained from the experiment done in this research; on varying amounts of distortions in ROI and non-ROI indicate that

it would be better to use more number of bits for the interesting regions in an image. Thus this concept can be used in compression to reduce the bit rate while retaining the quality of the image. This can be done by allocating very little bandwidth for unimportant regions in an image like the background. Digital watermarking is a pattern of bits inserted into a digital image that identifies the file's copyright information. The purpose of digital watermarks is to provide copyright protection for intellectual property that's in digital format. The actual bits representing the watermark must be scattered throughout the file in such a way that they cannot be identified and manipulated. But for better results it is suggested that these bits be placed in regions of the image like the background. Thus placing these bits along with the bit stream of the ROI should be avoided.

This concept can also be used for unequal error protection. Most of the error protection techniques that are presently being used have equal error protection. That is all the bits in a bit have equal error protection. In recent times studies have showed that some regions in the image need to have better error protection compared to other regions. This stems from the fact that some regions have to be error free at any cost. Apart from this there is also a possibility of reduced computation wherein only the important regions are processed.

1.1 Motivation

Inspite of having so many applications in various fields and the extensive research work going on in this field there have been very few image processing algorithms that could identify the ROI or predict an importance map accurately. A number of attempts have been made in this direction. Some of them use top-down approach while others use

bottom-up approach. But there still exists a need for an automated image processing algorithm that can predict the importance map for any given image.

Thus an attempt has been made in this research to establish the importance of identifying the ROI and an algorithm is proposed which would predict an importance map. Thus in this paper, I present research toward:

1) Providing quantitative evidence that knowledge of ROI has an effect on the image quality. The effect of ROI on image quality in different surrounding is examined.
2) Quantifying the relationship between measurable factors and perceived interest of each of these.
3) Incorporate the above results into a semi-automated algorithm for determining the importance map for a given gray scale image.

The following section reviews previous attempts at identifying ROI's.

1.2 Previous work

There has been a lot of research going on in this field. Various attempts have been made to establish the importance of defining ROI in an image. Numerous experiments were performed to testify that ROI coding could be used to improve image quality. This presents the strong need for an automated image processing algorithm. Some of the important works in this field are mentioned below.

The use of edge detectors was considered but the results were not satisfactory. Figure 1.1 shows the output of an edge detector.

(a) *(b)*

***Figure 1.1**: (a) Image of a horse (b) Output of the edge detector*

Another obvious technique at identifying ROI's is to use average region brightness. That is the image is divided into various regions and the average brightness value of each of the regions is calculated. The importance map is then constructed based on its importance value. One such importance map is shown below

(a) *(b)*

***Figure 1.2**: (a) Input image (b) Output importance map based on brightness*

Andrew Bradley [1] in his paper investigates the possibility of improving the overall perception of image quality by preferentially coding certain regions of interest (ROI) in an image. Experiments were conducted utilizing an automated algorithm for visual attention to detect the primary ROI(s) in an image, and then encoding the image using the maxshift algorithm of JPEG 2000. The results indicate that, while there is no overall preference for the ROI encoded images, there is an improvement in perceived image quality at low bit rates (below 0.25 bits per pixel). Liu and Fan [4] in their paper propose a new ROI coding method called Partial Significant Bit planes Shift (PSBShift) that combines the advantages of the two standard ROI coding methods defined in JPEG2000. The PSBShift method not only supports arbitrarily shaped ROI coding without coding the shape, but also enables the flexible adjustment of compression quality in ROI and background.

Privitera and Stark [5] use Eye-fixations to determine the human defined ROIs. They also use some image processing algorithms to find the algorithmically defined ROIs. These hROIs and aROIs are then compared through analysis of their spatial locations and analysis of the temporal order or sequential binding. These results measure the capability of image processing algorithms in finding the ROI's. Nguyen and Chandran [6] analyze the spatial and temporal characteristics of the human visual attention system as recorded from an eye-tracking device at the encoding end. They establish that human visual attention mechanisms direct the viewer's eye movements around the image to provide a sequence of fixations. These fixations are analyzed,

clustered and classified into regions of interest (ROI). These ROIs are used to selectively encode and prioritize regions such that an improved image content recognition performance can be achieved.

Marichal and Delmot [8] introduce a tool that attempts to find which regions are important in an image or in a video sequence. For this purpose, Fuzzy was used to modelize human Subjective knowledge. The resulting classification was used in a wide range of applications going from image coding to image quality. Peli [7] proposes an algorithm for the detection of visually relevant luminance features. The algorithm detects edges (sharp luminance transitions) and narrow bars (luminance cusps) and marks them with the proper polarity. The algorithm was robust with respect to variations in filter parameters and requires no use of quadrature filters or Hilbert transforms.

Laurent Itti [10] proposes a neuro-behavioral model which takes into account bottom-up saliency and does not require any top-down guidance to shift attention. In this approach the input is decomposed into a set of topographic feature maps based on its color, intensity and orientation. Different regions then compete for saliency within a map such that only those standing out from the others get a higher priority. All these different maps are then combined into a single master ‘Saliency map’ detecting the important regions.

When Itti’s code is applied to an image the output is as shown below.

(a) *(b)*

***Figure 1.3**: (a) Dog image (b) Output of Itti's algorithm*

Osberger and Maeder [9] propose a model based on human visual attention and eye characteristics. Several features that influence human visual attention are evaluated for each region of a segmented image. These are then combined to produce an importance map that shows the importance of each region. They investigate different features like Contrast, Size, Shape, Location, Color and Motion. Equations are then postulated to give the importance of each of these factors. All these factors are then combined to form an importance map. In spite of their results being good, one of their limitations is that there is no computational or experimental method for arriving at these equations. The algorithm also uses heuristic equations that have not been experimentally verified in terms of human vision. It also combines the various factors using simple sum of squares. Thus it is assumed that all of these factors are of equal importance. This is a

potential problem because each factor has its own importance. The results obtained by using their algorithm are shown below:

(a) *(b)*

Figure 1.4: (a) Image of man and dog (b) Output of Osberger's algorithm

Thus as seen from the above results there still exists a need for an automated image processing algorithm that can predict the importance map for any given image.

One goal of this research is to use human psychophysics experiments to quantify the relationship between different factors like size, location, contrast, blur, category and the perceived importance. And a second goal of this work is to determine the best way to combine the individual perceived importance's to arrive at the overall perceived importance of each region

1.3 Outline

The thesis has been organized into five different chapters. The first chapter consists of this introduction and it highlights the motivations and contributions of this work. The second chapter consists of the experiment done to provide quantitative evidence that knowledge of ROI has an effect on the image quality. The importance of defining ROI in an image is clearly established in this experiment. Chapter three presents the second experiment performed to quantify the relationship between measurable factors that determine ROI and perceived interest of each of these. This explains in detail the importance of each of these factors in determining the ROI. Chapter four presents an algorithm to automatically predict an importance map for any given segmented image. An optimization is run on different factors to arrive at the best possible solution. Finally chapter five contains the conclusion and ideas for future work along the same line.

CHAPTER 2

Experiment-1

The main aim of this experiment was to quantitatively show that, the knowledge of ROI in a given image would have a considerable effect on image quality. The study is also aimed at showing that when this knowledge is applied during image compression there can be considerable improvement in image quality at almost the same bit-rate. The presence of persons, animals, objects in an image generally determine the important regions in an image and its effects are studied briefly. In this experiment, it is examined whether an image's perceived fidelity is affected by the perceived fidelity of the regions of interest (ROIs) to a greater extent than it is by the perceived fidelity in the other regions. It is generally expected that subjects would prefer images containing most of the distortions in the non-ROI. However, we also expect that, at some point, the non-ROI becomes so distorted that the overall perception of fidelity is reduced. Thus this experiment was also aimed at predicting a good way of proportioning the distortion between the ROIs and the non-ROI. Insights into this issue are particularly important for the design of image processing systems which possess some control over the amount of distortion that is induced in different regions of images.

Eleven images were distorted in one of three ways (white noise, blurring, and wavelet subband quantization distortion), and then ratings of perceived fidelity were obtained for the distorted images.

2.1 Methods

2.1.1. Apparatus

Stimuli were displayed on a high-resolution, Proview PS-910 19-inch monitor (Proview Technology, Inc., Garden Grove, CA, USA) with a 0.25 mm dot pitch and maximum horizontal and vertical scan frequencies of 85 kHz and 150 Hz, respectively. The display was operated at a resolution of 35.1 pixels/cm and a frame rate of 75 Hz. The display yielded minimum and maximum luminances of, respectively, 0.55 and 100.0 cd/m2, and an overall gamma of 2.6. Luminance measurements were made by using a Minolta CS-100A photometer (Minolta Corporation, Tokyo, Japan). Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 46 cm resulting in a display visual resolution of 28.4 pixels/deg.

2.1.2. Stimuli

Stimuli used in this study were 512 × 512-pixel grayscale images which subtended 18 × 18 degrees. The images were cropped from 768×512-pixel originals obtained from the McGill calibrated image database. Eleven original images were selected: Four of the images contained frontal/side views of human faces (kids, cyclists, students, man); three of the images contained animals (bird, dog, duck); and four of the images contained man-made objects or a combination of human/animal/plant and man-made objects (boat, hydrant, flower, cityscape). The 11 original 8-bits/pixel RGB images were gamma-corrected for the acquisition process such that digital pixel values were proportional to luminances in the physical scenes (within limitations of the acquisition devices). [35]

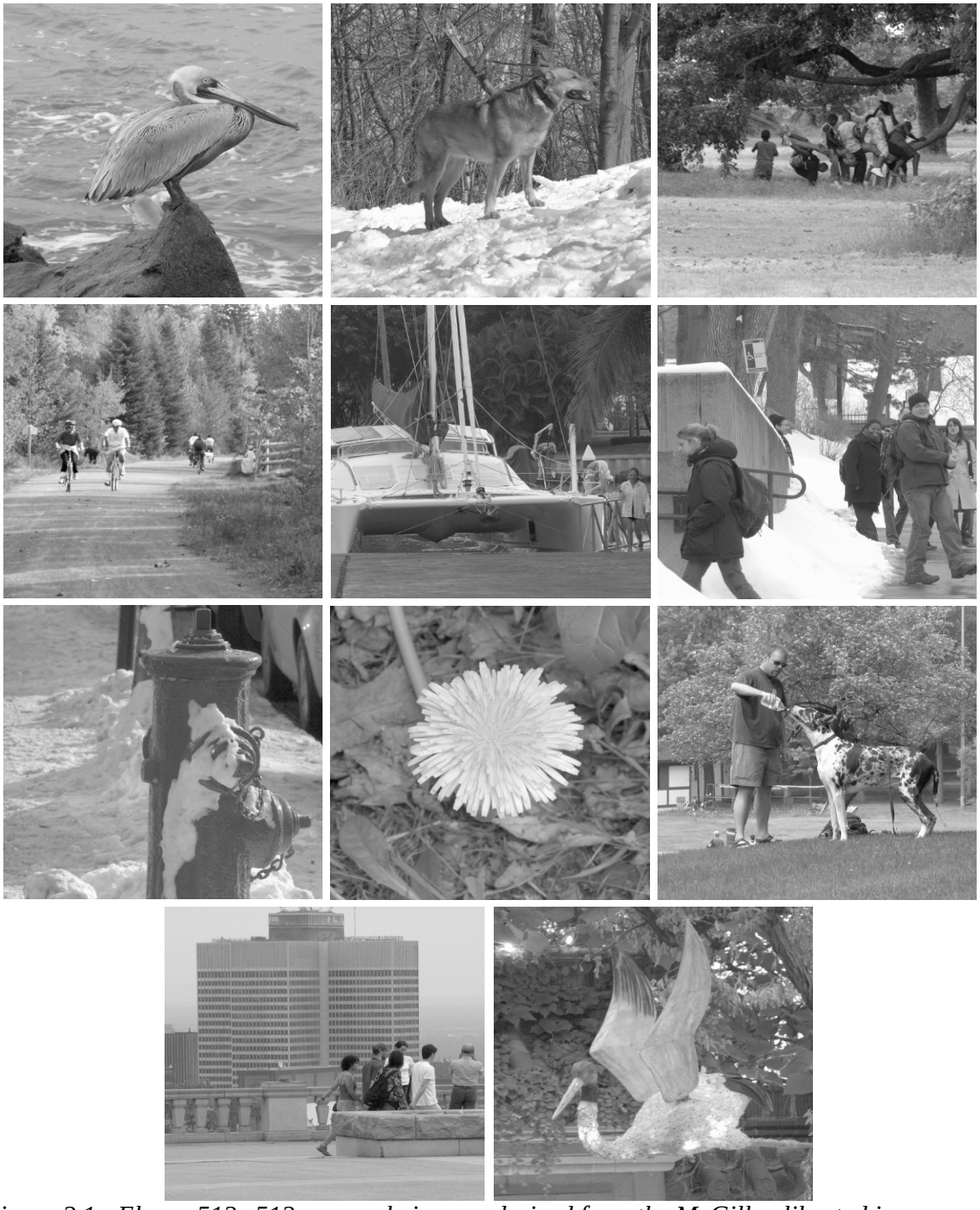

Figure 2.1: Eleven 512×512 grayscale images derived from the McGill calibrated image database. From top to bottom, left to right, the common names that have been assigned to these images are: bird, dog, kids, cyclists, boat, students, hydrant, flower, man, people, and duck.

The images were then converted to grayscale via a pixel-wise transformation of I = 0.2989R + 0.5870G + 0.1140B, where I, R, G, and B denote the 8-bit grayscale, red, green, and blue intensities, respectively. The grayscale values were then transformed via $I' = I^{1/\gamma}, \gamma = 2.6$ so as to maintain the linear relationship between original pixel values and luminance when the images were shown on the display apparatus. These modified grayscale values were then scaled to span the range 0 − 255.

A section of size 512×512 pixels was then cropped from each image so as to capture at least one region of interest. For each 512 × 512 grayscale image, one or more ROIs were manually selected by following these criteria: Regions of high contrast; larger regions; objects in the foreground; regions containing plants, animals, and humans.

A 512 × 512 binary mask was then created in which pixels with a value of 1 corresponded to the selected ROIs. Figure 2.1 depicts the 512 × 512 images used in this study; Figure 2.2 depicts the corresponding masks.

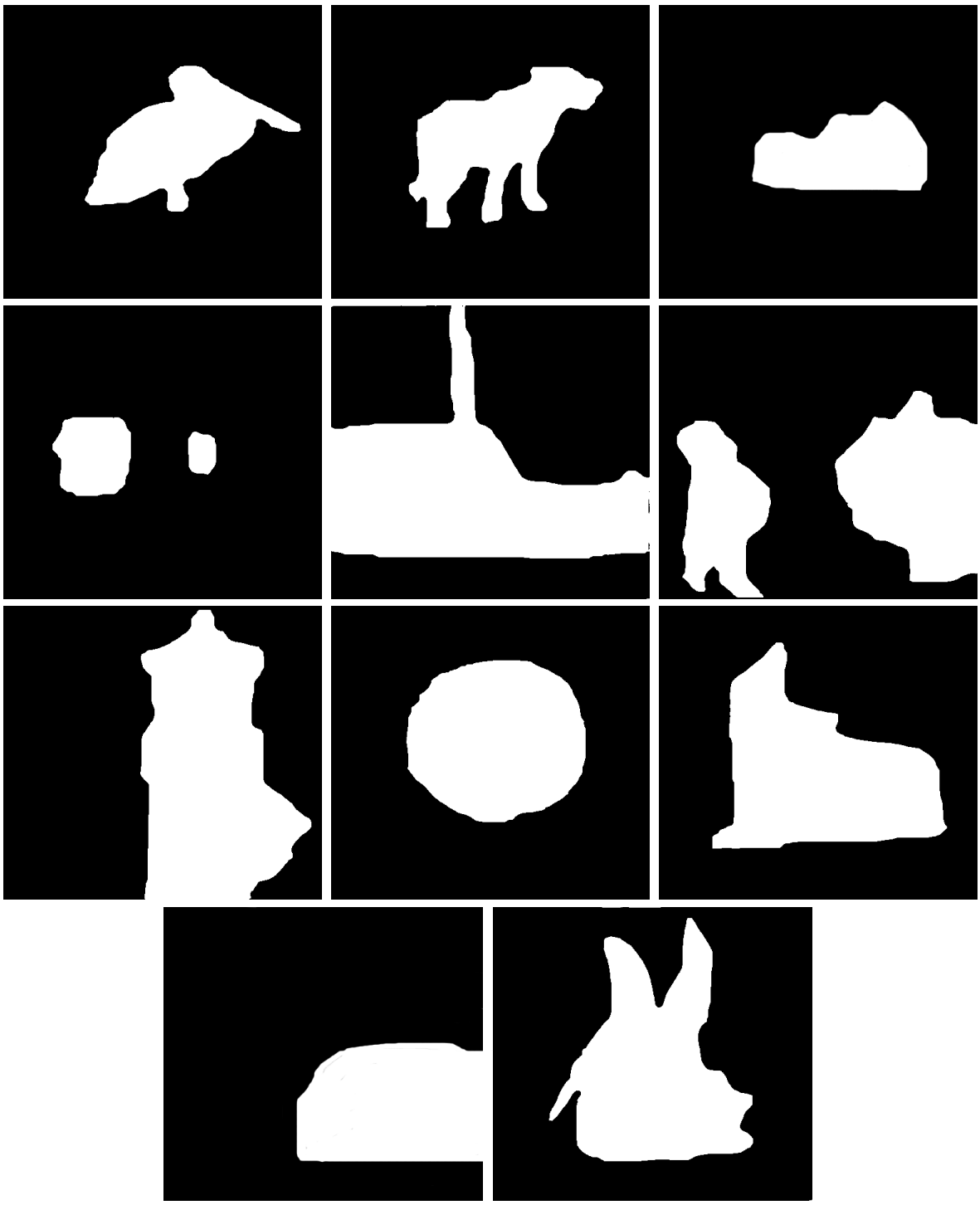

Figure 2.2: Binary masks created by the author corresponding to each of the images in Figure 2.1. White pixels denote to the assumed ROI selected by author; black pixels denote the non-ROI. From top to bottom, left to right, these masks correspond to image: bird, dog, kids, cyclists, boat, students, hydrant, flower, man, people, and duck.

Each original 512 × 512 grayscale image was distorted in three ways:

1. White noise in which the pixel values were drawn from a zero-mean Gaussian distribution with standard deviation adjusted to achieve the desired distortion contrasts.
2. Blurring by using a length-37 finite impulse response filter with filter coefficients selected according the Gaussian function

$$h(n) = \exp(\frac{-(n/18.5)^2}{2\sigma^2}), n \in [-18,18],$$

with σ selected to achieve the desired distortion contrasts; this one-dimensional filter was first applied to the rows of each image and then to the columns of that result to generate the final blurred image.

3. Wavelet subband quantization distortion induced by (a) performing a 5-level discrete wavelet transformation of the image by using the 9/7 bi-orthogonal filters; then (b) uniformly quantizing the coefficients within each subband such that the mean squared error (MSE) within each subband was proportional to $2^{-n} D_{base}$, where n denotes the decomposition level and D_{base} denotes the baseline MSE selected to achieve the desired distortion contrasts; and then(c) performing an inverse DWT to generate the distorted image.

The distortions were added separately to the ROI and non-ROI by using each image's corresponding mask. Specifically, let $\hat{I}_{ROI}$ denote a distorted version of original image I such that the RMS contrast of the distortion in the ROI is C_{ROI}. Let $\hat{I}_{non-ROI}$ denote a distorted version of I such that the RMS contrast of the distortion in the non-ROI

is $C_{non-ROI}$. Let M denote the mask with pixel values in the range 0−1, where pixel values of 1 correspond to the ROI(s) and pixel values of 0 correspond to the non-ROI. The total distorted image $\hat{I}$ containing distortion at contrast C_{ROI} in the ROI and distortion at contrast $C_{non-ROI}$ in the non-ROI was generated via

$$\hat{I} = \mathrm{M} \times \hat{I}_{ROI} + (1 - \mathrm{M}) \times \hat{I}_{non-ROI} . \qquad (1)$$

The distortions were generated such that the RMS contrast of the distortions over the entire image was 0.15. The RMS distortion contrast in the ROI, CROI, was set to fixed values of either 0 (no ROI distortion) or $C_{ROI,\max}$ (all distortion in the ROI), where $C_{ROI,\max}$ denotes an image-specific value required to achieve a total RMS distortion contrast of 0.15. Six intermediate values of C_{ROI} were then selected according to a logarithmic spacing to fall between these two extremes: Three intermediate values of C_{ROI} were chosen to conform to a logarithmic spacing between [0.05, 0.15) and the remaining three intermediate values were chosen to conform to a logarithmic spacing between $[0.15, C_{ROI,\max}]$.

For distortions consisting of white noise and blurring, the desired distortion contrasts were met by adjusting the standard deviation of the underlying Gaussian. For the wavelet subband quantization distortion, the contrasts were met by adjusting the baseline MSE Dbase and thereby adjusting the granularity of the quantizer. Thus, a total of 588 images were used in Experiment I: Eleven original images and 264 distorted images (11 images × 3 distortion types × 8 ROI/non-ROI proportions).

Figure 2.3 depicts representative distorted images used in the experiment containing white noise, blurring, or wavelet subband quantization distortion.

Figure 2.3: Representative distorted images used in Experiment 1. ***Top row****: Image bird containing white noise.* ***Middle row****: Image kids containing blurring****. Bottom row****: Image flower containing wavelet sub band quantization distortions. Images in the first column have all distortion in the non-ROI (no distortion in the ROI). Images in the second column have equal distortion contrast in the ROI and non-ROI. Images in the third column have all distortion in the ROI (no distortion in the non-ROI).*

2.1.3. Procedures

Subjective ratings of fidelity were measured for each distorted image by using a modified version of the Subjective Assessment Methodology for Video Quality (SAMVIQ) testing procedure applied to still images. Each experimental session began with three minutes each of dark adaptation and adaptation to a blank 19.0 cd/m^2 display. Subjects were then concurrently shown an original image and one of the eight distorted versions of that image; each of the eight distorted images contained the same type of distortion at a different ROI distortion contrast. The original image and the currently-visible distorted image were placed upon a uniform 19.0 cd/m^2 background and horizontally separated by approximately 5 degrees (measured from the right edge of one image the left edge of the other); the original image was always located in the left-hand position. Figure 2.4 shows a screen shot of the experimental setup.

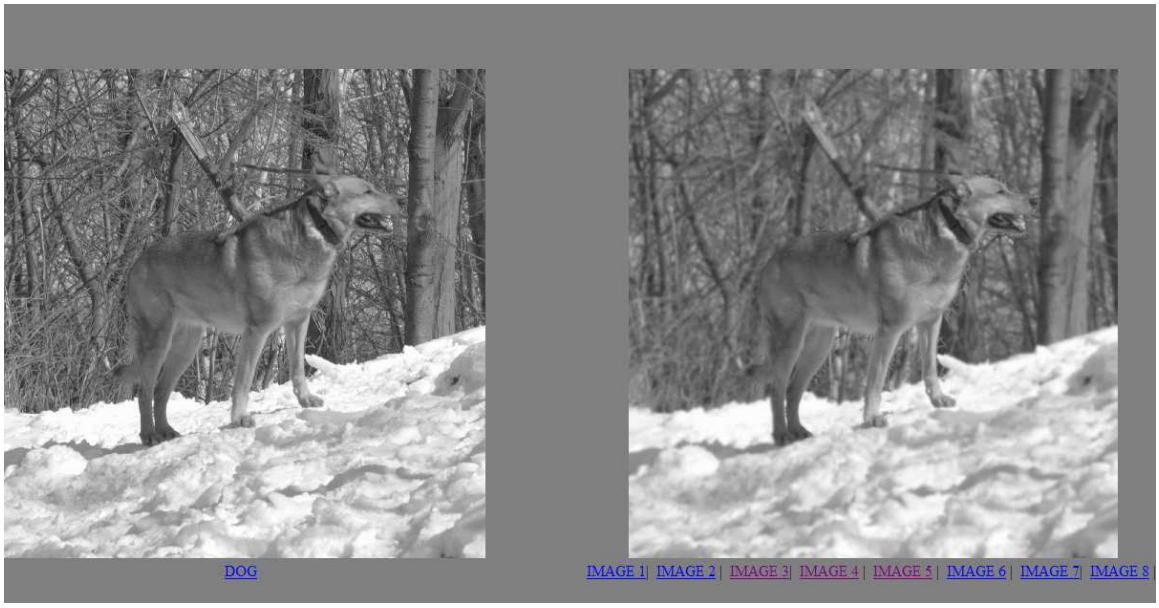

Figure 2.4: Screenshot of experimental set-up

Via keyboard input, subjects switched between the eight distorted images. Before a new distorted image was shown, a zero-contrast, 19.0 cd/m^2 image was displayed for 750 ms so as to prevent apparent motion of the distortion. Subjects were instructed to provide for each distorted image a rating of perceived quality relative to the original on a scale from 0 − 100 (where 100 denoted a quality equal to the original). Subjects were free to switch between any of the eight distorted images and to change any previously reported ratings for those distorted images. No specialized instructions were given to the subjects regarding where in the images to look; subjects were instructed only to

1) Provide a rating of quality relative to the original on a scale from 0−100,
2) Review and confirm their results for all eight images before moving on to the next set of images.

After rating all eight distorted images corresponding to a particular original image, the experimental session continued with a new original image and eight distorted versions of the new original image. Ratings were reported verbally and recorded by a proctor.

Each of the eight distorted images contained distortions with an RMS contrast in the ROI, C_{ROI}, corresponding to values in the range $[0, C_{ROI,\max}]$ as described. All distorted images for a particular experimental session contained the same type of distortion (noise, blur, or wavelet compression). Thus, a single experimental session consisted of rating 88 images (8 distorted versions of 11 original images) containing the same type of distortion. The entire experiment consisted of three experimental sessions, one for each type of distortion. The time-course of each experimental session was not

limited; however the majority of observers completed each session (rating 88 images) in approximately 35 minutes. The total time required to complete all three experimental sessions varied across subjects; however, all results were obtained within the course of three weeks. We acknowledge that adaptation had likely occurred during the course of each session, and that learning may likely have occurred over the course of the experiment. These are largely unavoidable secondary effects of the SAMVIQ paradigm. However, unlike pair wise comparison procedures which require remembering previous ratings, the paradigm employed here allowed subjects to account for any bias induced by previous judgments. Subjects frequently made adjustments to previous ratings, and all subjects performed a final pass on each set of images to ensure their satisfaction in their ratings.

2.1.4. Subjects

The author, two adult image-processing researchers familiar with the purpose of the study, and three naive adult subjects participated in the experiments. The image-processing researchers had previously encountered the types of distortions used in this study (noise, blurring, and wavelet compression); however, only the author had previous exposure to the images. Subjects ranged in age from 22 to 31 years. All subjects had either normal or corrected-to-normal visual acuity.

2.1.5 Contrast Metric

Results are reported here in terms of RMS contrast, which is defined as the standard deviation of luminances normalized by the mean luminance of the background. RMS contrast has been applied to a variety of stimuli, including noise, wavelet, and natural images. In this thesis, results are reported in terms of the RMS contrast of the distortions computed with respect to the mean luminance of the original image. The RMS contrast of the distortions, Crms , is given by

$$C_{rms} = \frac{1}{\mu_{L(I)}}\left(\frac{1}{R}\sum_{i=1}^{R}[L(E_i) - \mu_{L(E)}]^2\right)^{\frac{1}{2}} \qquad 2$$

where R denotes the number of pixels in the region over which the distortion contrast is measured; where $\mu_I = \frac{1}{N}\sum_{i=1}^{N} I_i$ and $\mu_{L(I)} = \frac{1}{N}\sum_{i=1}^{N} L(I_i)$ denote the average pixel value and average luminance of I, respectively, and N denotes the number of pixels in I; and where $\mu_{L(E)} = \frac{1}{R}\sum_{i=1}^{R} L(E_i)$ denotes the average luminance of the mean-offset distortions $E = \hat{I} - I + \mu I$. The quantities L(Ii) and L(Ei) correspond to the luminance of the ith pixel of the image and the mean-offset distortions, respectively.

2.2. Results and Analysis

The eight ratings of perceived fidelity obtained for each observer on each image set (image/distortion-type combination) were converted to z-scores such that the ratings for each observer on each image set had a zero mean and unit standard deviation. Accordingly, these standardized subjective ratings are interpreted as relative values of perceived fidelity. The graphs are then drawn for change in perceived interest with respect to relative distortions. The graphs for all the different distortion types are shown below.

Noise (high) refers to the high contrast noise condition wherein the noise is added to the image in such a way that the RMS contrast is 0.15. Similarly Noise (low) refers to the low contrast noise condition wherein the noise is added to the image in such a way that the RMS contrast is 0.75.

Wavelet (high) refers to the high contrast wavelet distortion wherein the distortion is added to the image in such a way that the RMS contrast is 0.15. Similarly Wavelet (low) refers to the low contrast wavelet distortion wherein the distortion is added to the image in such a way that the RMS contrast is 0.75.

Blur (high) refers to the high contrast blur condition wherein the Gaussian blur is added to the image in such a way that the RMS contrast is 0.15. Similarly Blur (low) refers to the low contrast blur condition wherein the Gaussian blur is added to the image in such a way that the RMS contrast is 0.75.

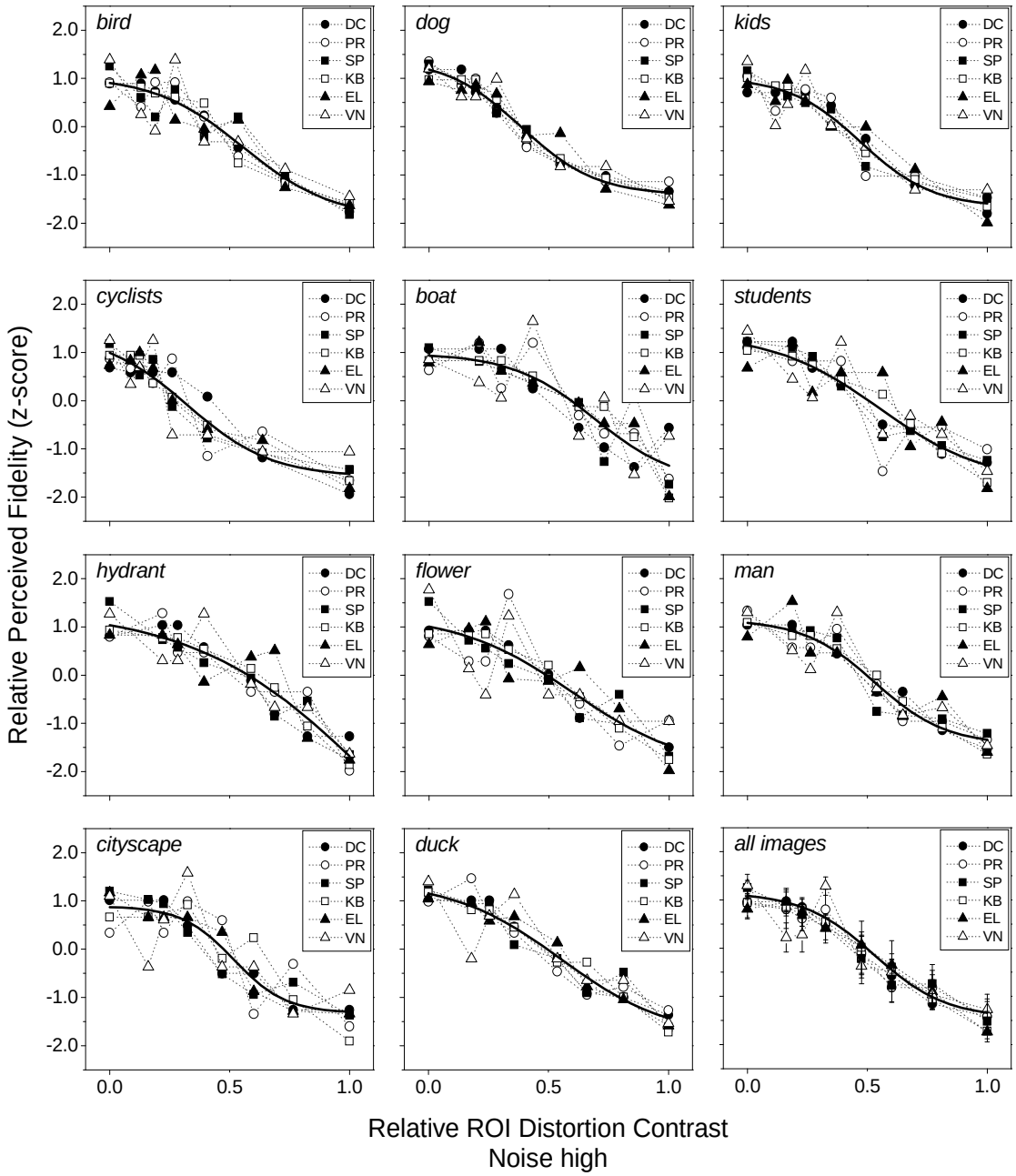

Figure 2.5: Subjective ratings of fidelity for images containing Gaussian white noise (High).

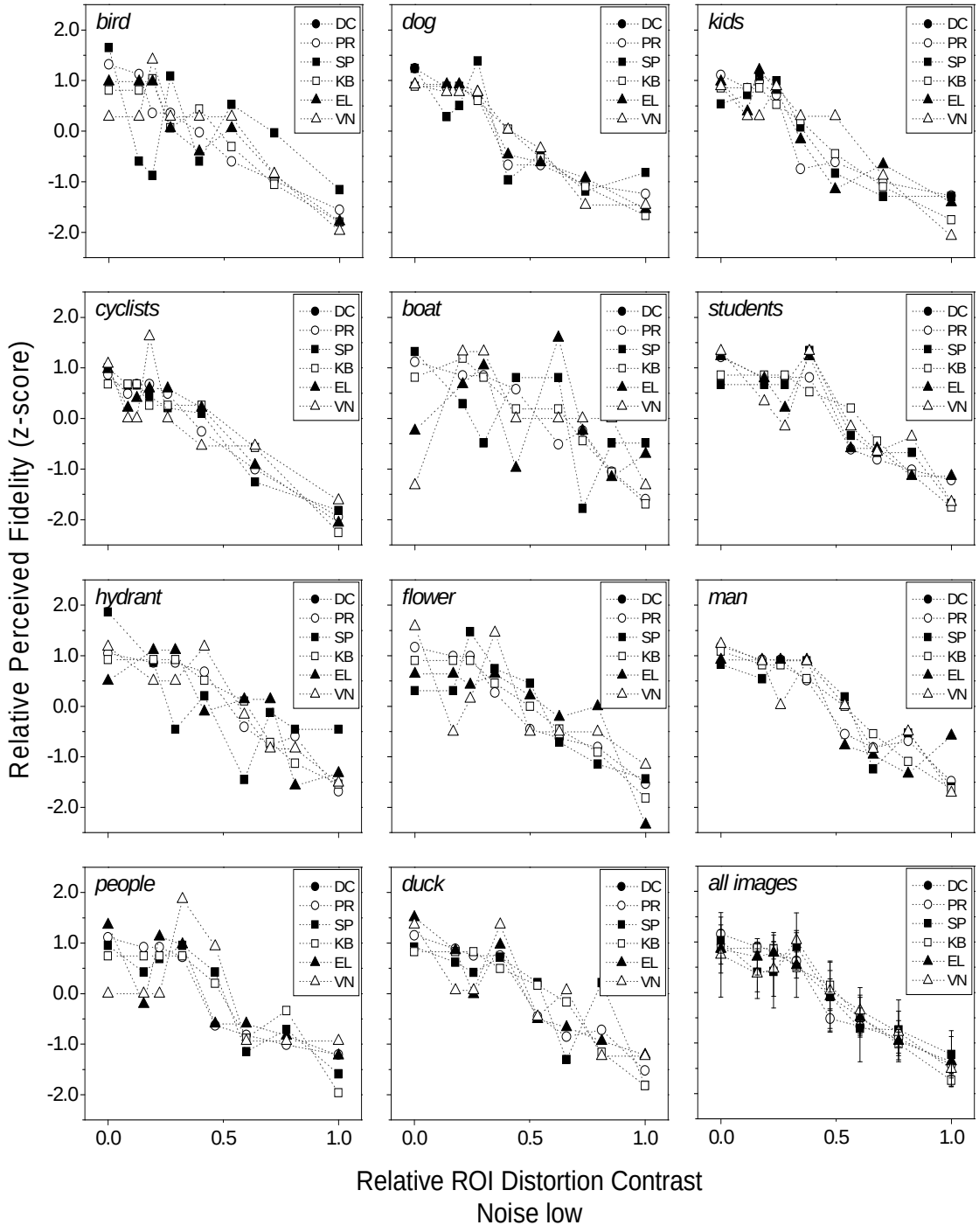

Figure 2.6: Subjective ratings of fidelity for images containing Gaussian white noise (Low)

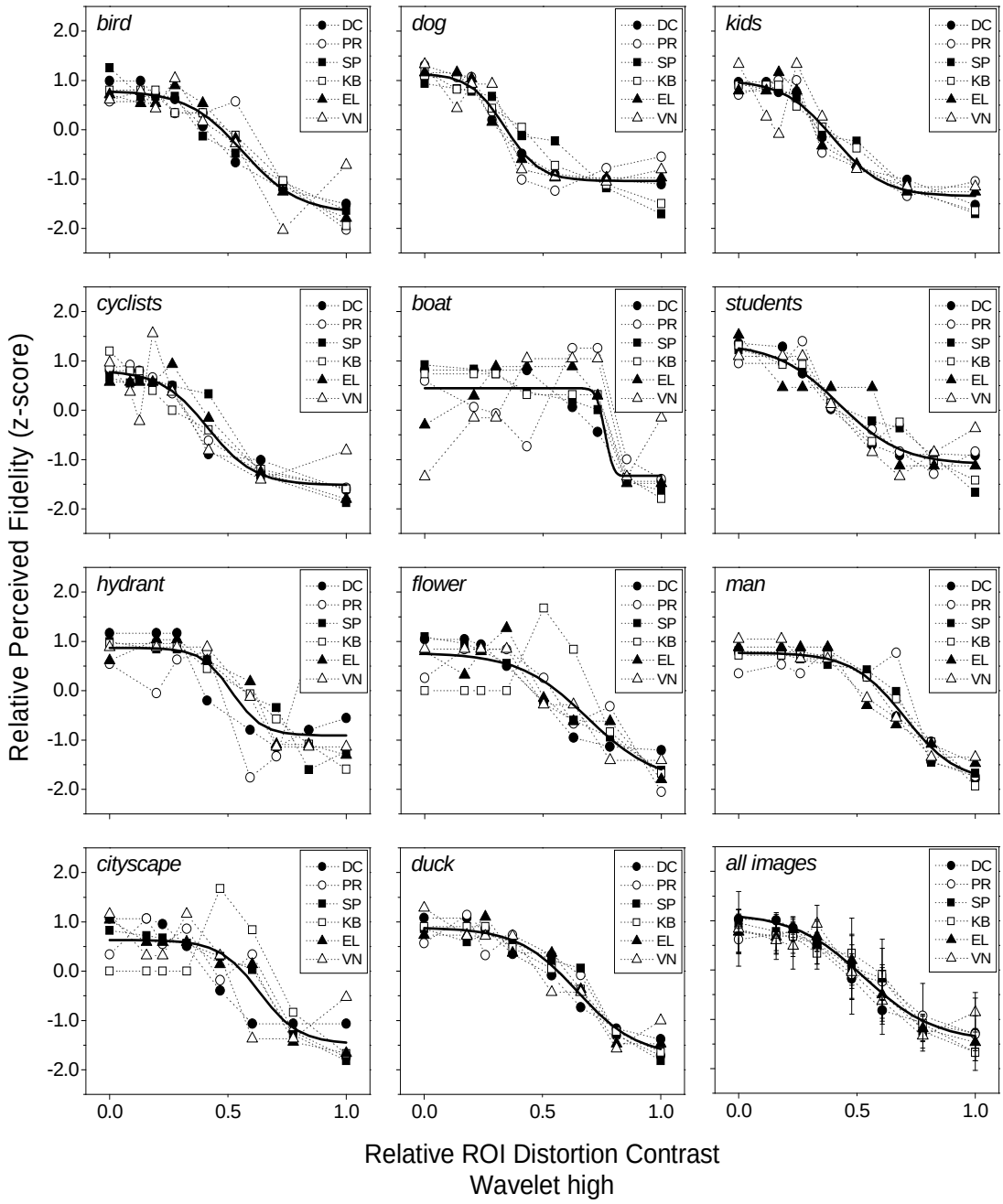

Figure 2.7: Subjective ratings of fidelity for images containing wavelet subband quantization distortion (High)

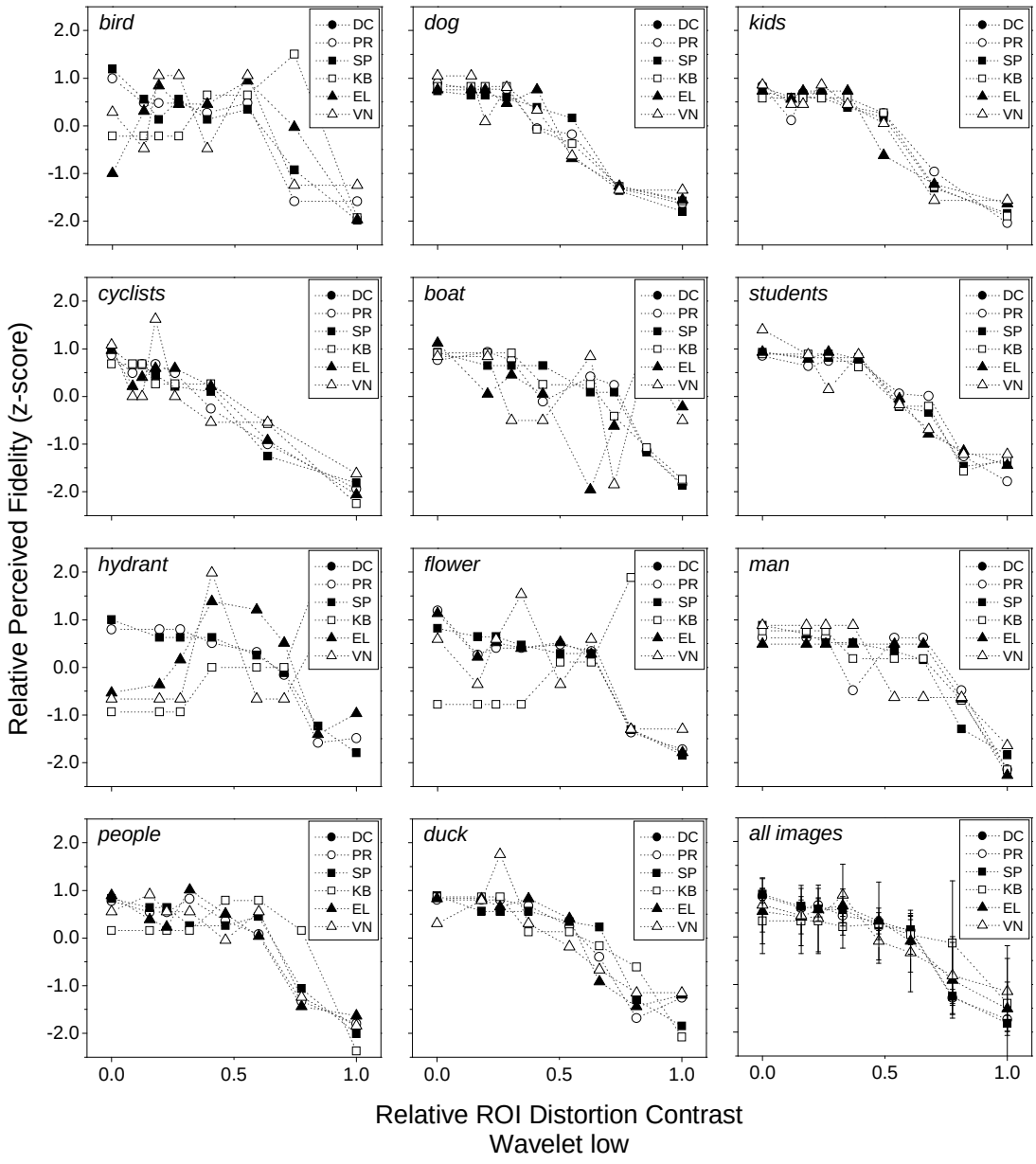

Figure 2.8: Subjective ratings of fidelity for images containing wavelet subband quantization distortion (Low)

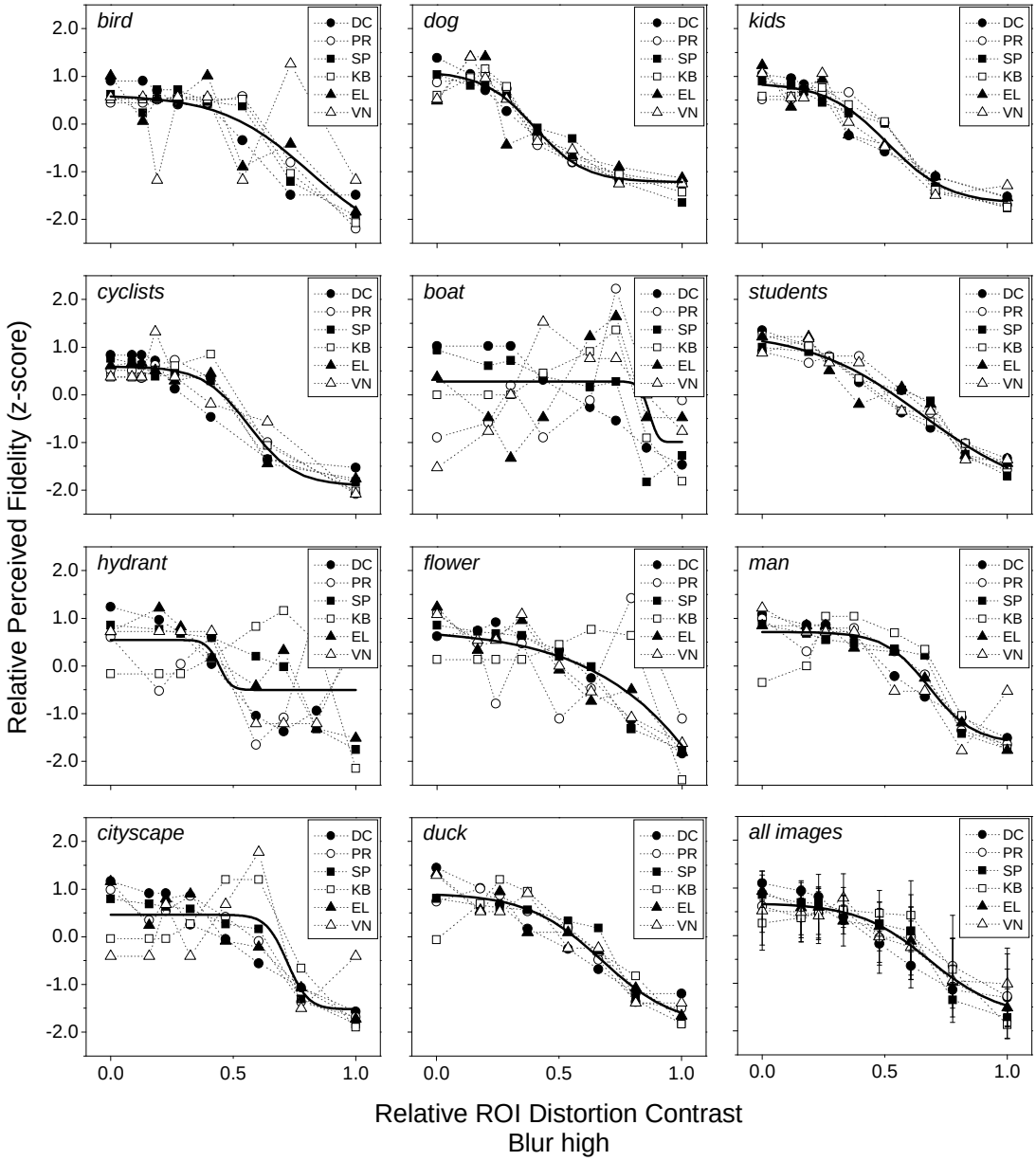

Figure 2.9: Subjective ratings of fidelity for images containing Gaussian blurring (High)

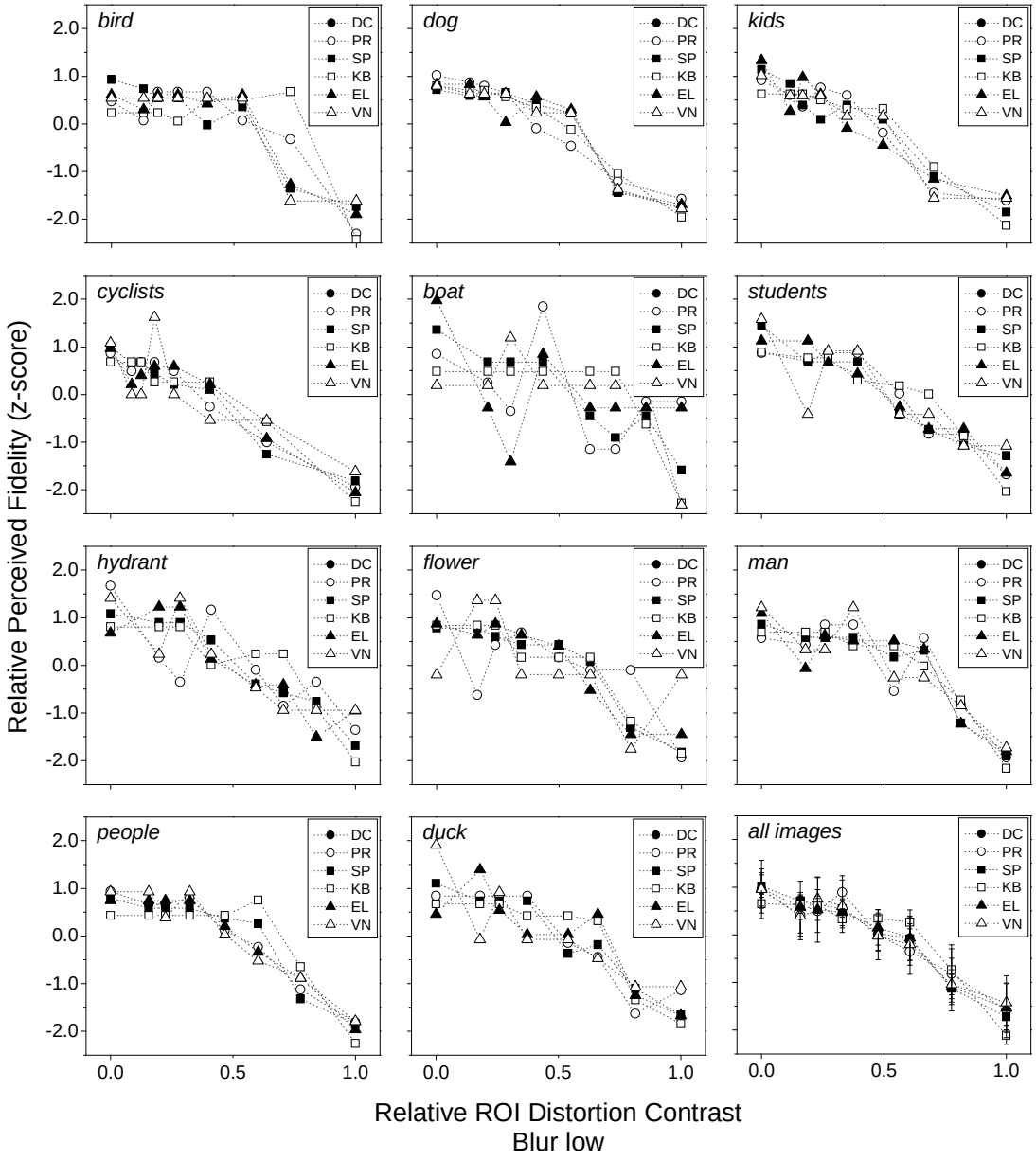

Figure 2.10: Subjective ratings of fidelity for images containing Gaussian blurring (Low)

Figures 2.5 – 2.10 depict results obtained for images containing white noise, wavelet subband quantization distortion and blurring, respectively.

Each graph in Figures 2.5 – 2.10 depict results obtained for individual images. The horizontal axis denotes the relative RMS distortion contrast in the ROI (the ratio of the distortion contrast in the ROI to $C_{ROI,\max}$). The vertical axis denotes relative perceived fidelity (z-score). Individual data points correspond to a particular rating from a particular subject; filled symbols correspond to results obtained from the subjects who were familiar with the purpose the experiment (the author); open symbols correspond to results obtained from the naive subjects. The solid line in each graph denotes the best-fitting sigmoid function, provided to help visualize general trends. The graph in the lower right-hand corner of each figure depicts results averaged over all 11 images for each subject. Although there is clearly variability in the results obtained for different subjects on different images, several general observations can be drawn from these data.

First, observe from Figures 2.5 and 2.6 (results for white noise) that relative perceived fidelity tends to decrease as more of the distortion is moved from the non-ROI into the ROI. For the majority of images, perceived fidelity begins to demonstrate a marked drop as the relative ROI distortion contrast increases to approximately 0.4−0.5. In addition, relative ROI distortion contrasts below 0.4 tend to give rise to similar relative perceived fidelity, an effect which is likely attributable to masking. Furthermore, for the

majority of images, a relative ROI distortion contrast of 1.0 tends to give rise to the least (or amongst the least) relative perceived fidelity.

With notable exceptions, these trends are also suggested by the data in Figures 2.7-2.10 (results for blurring and wavelet subband quantization distortion, respectively). For the majority of images, perceived fidelity begins to demonstrate a marked drop as the relative ROI distortion contrast increases to approximately 0.4-0.5. The major exceptions to this finding appear to occur for images boat, hydrant, and cityscape, which contain a substantial amount of man-made structure, and in which the selected ROIs are not overwhelmingly more interesting than the selected non-ROIs. Notice from these data that images boat, hydrant, cityscape, and flower tend to demonstrate the most variability among subjects for all three types of distortion. Conversely, images which contain human or animal faces tend to demonstrate the least variability across subjects.

When the results for individual images are compared across distortion type, the results are substantially more image-specific. Observe from Figures 2.5-2.10 that images which contain animal and human faces demonstrate roughly the same general trend for all three types of distortion (e.g., dog, kids, students, man). Images which lack such features, e.g., boat and hydrant, exhibit marked distortion-type-specific trends. Also observe that white noise, which is uncorrelated with the image's edges, appears to induce less of an image-specific effect on perceived fidelity than blurring and wavelet subband quantization distortion, distortions which tend to degrade the image's edges.

Overall, the results of this experiment are consistent with the assertion that overall perceived fidelity is affected by what in the image is distorted. Here, we have shown this effect to occur for all three types of distortion: white noise, blurring, and wavelet subband quantization distortion. For the images tested here, our results suggest that for a fixed total RMS distortion contrast, a relative ROI distortion contrast of approximately 0.4−0.5 can be induced in the ROI without substantially affecting perceived fidelity.

However, our results have also demonstrated that the extent to which perceived fidelity is affect by ROI vs. non-ROI distortion is largely dependent on the extent to which the ROI is more interesting than the non-ROI. For all three types of distortion, images in which the ROIs were not overwhelmingly more interesting than the non-ROIs, demonstrated the greatest variability in perceived fidelity across subjects; whereas images in which the ROIs were substantially more interesting than the non-ROIs (e.g., images containing faces) demonstrated the least variability across subjects.

CHAPTER 3

Experiment – 2

In this experiment an attempt has been made to estimate the importance of various factors in determining the ROI for a given image. We examined from a psychological standpoint how each of these factors was influencing the ROI. Different factors like Size, Location, Blur and Contrast were tested in our experiment. Here an assumption was made that all these factors are independent of each other, so each of these factors was tested independently. A set of images were then shown to different subjects to get their objective ratings. Based on these ratings importance graphs for each of them were predicted.

3.1 Methods

3.1.1. Apparatus

Stimuli were displayed on a high-resolution, View sonic VA912b 19-inch monitor with a 0.25 mm dot pitch and maximum horizontal and vertical scan frequencies of 85 kHz and 150 Hz, respectively. The display yielded minimum and maximum luminance of respectively, 2.7 and 207.0 cd/m^2, and an overall gamma of 2.9. Luminance measurements were made by using a Minolta CS-100A photometer (Minolta Corporation, Tokyo, Japan). Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 46 cm resulting in a display visual resolution of 28.4 pixels/deg.

3.1.2. Stimuli

A single grayscale natural scene obtained from the van Hateren database [37] served as a common background for all stimuli. The original image was of size 1536 × 1024 pixels with 16-bit pixel values in which each pixel value was proportional to luminance in the original physical scene. This image was modified by (1) resizing the image to 1024×768 pixels; then (2) applying a point-wise power function of $f(x) = x^{1/2.9}$ (where x denotes the original pixel value) such that the displayed pixels were proportional to luminance in the original physical scene; and then (3) scaling the pixel values to lie in the range [0, 255].

Figure 3.1: Background from van Hateren Database

Three additional high-resolution single-object images were obtained from the Microsoft ClipArt collection [38] to serve as the regions (objects) of interest. These objects consisted of an image of a human (image *soldier)*, an image of an animal (image *dog*), and an image of a non-human/non-animal object (image *hydrant)*. The original 24-bits/pixel color images were converted to 8-bit grayscale, and then a mask was drawn such that only the pixels within each object were non-transparent.

(a) *(b)* *(c)*

Figure 3.2*: (a) Dog (b) Soldier (c) Hydrant*

To each of the three single-object images, the following manipulations were applied:

1. *Size*: The single-object images were resized (via bi-cubic interpolation) such that the number of pixels in the object were 1%, 2%, 3%, 6%, and 12% of the number of pixels in the natural-scene background. These objects were then placed within the natural-scene background centered horizontally and displaced vertically so as to create a natural impression of depth. Figure 3.3 shows an example of variations in size.

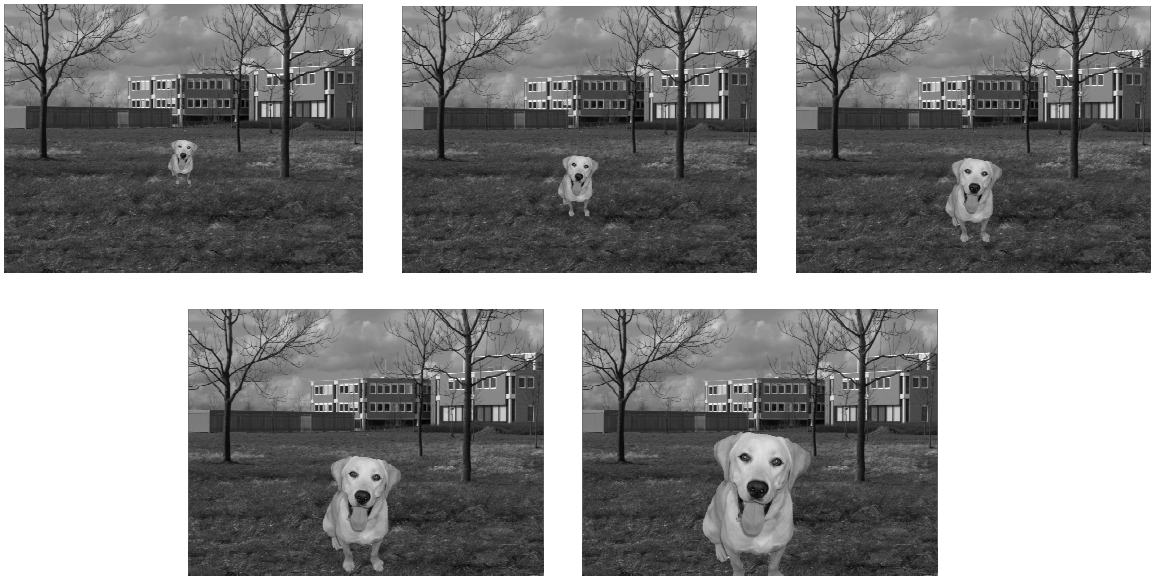

Figure 3.3: *Five images showing variation in size*

2. *Location*: The objects were placed within the natural-scene background at horizontal offsets from the center of the image of 0%, 20%, 39%, 59%, and 78% of the natural-scene's half-width (512 pixels). Here, the size was held constant at 3% of the number of pixels in the natural-scene background. Figure 3.4 shows an example of variations in location.

Figure 3.4: *Five images showing variation in location*

3. *Blur:* The objects were blurred by using a length-15 Gaussian filter with one-dimensional impulse response $h(n) = \exp(\frac{-(n/7.5)^2}{2\sigma^2}), n \in [-7,7]$, with σ values of 0 (no blurring), 0.14, 0.29, 0.5, and 1.0; the blurred objects were then placed within the natural-scene background. Here, the size was held constant at 3% of the number of pixels in the natural-scene background, and the location was held constant at 59% of the natural-scene's half-width. Figure 3.5 shows an example of variations in blur.

Figure 3.5: *Five images showing variation in blur*

4. *Contrast*: The RMS contrast of the objects was adjusted to values of 0.1, 0.225, 0.35, 0.475, and 0.6; the contrast-adjusted objects were then placed within the natural-scene background. Here, the size was held constant at 3% of the number of pixels in the natural-scene background, the location was held constant at 59% of the natural-scene's half-width, and the objects were not blurred. Figure 3.6 shows an example of variations in contrast.

Figure 3.6: *Five images showing variation in contrast*

3.1.3. Procedure

Subjective ratings were measured for each set of stimuli by using a modified version of the Subjective Assessment Methodology for Video Quality (SAMVIQ) testing procedure applied to still images. The experiment was divided into four sessions: One session for each of the four factors (size, location, blur, contrast). Each session was further divided into three sub-sessions: One sub-session for each object type (man, dog, hydrant). Thus, each sub-session entailed rating the five images consisting of a single object which varied over the five values of the corresponding factor. Figure 3.7 shows a screen shot of the experimental setup.

***Figure 3.7**: Screenshot of experimental set-up*

During each sub-session, subjects were initially shown one of the five images. Via keyboard input, subjects switched between the five images and were instructed to provide for each object (man, dog, or hydrant) a rating of interest for the object on a scale from 0-100 where 100 corresponded to the greatest possible interest; the natural-scene background was not judged in this experiment. Subjects were instructed to view all five images before reporting any ratings, and subjects were also allowed to change any previously reported ratings during the course of each sub-session. The time-course of each experimental session was not limited, however the majority of observers completed all four experimental sessions in under 60 minutes.

3.1.4. Subjects

A total of five different subjects were used. The first two subjects were researchers in image processing. The other three were master's students naïve to the purpose of experiment. Subjects ranged in age from 21 to 28 years. All subjects had either normal or corrected-to-normal visual acuity.

Results

The raw scores for each subject on each set of five images (corresponding to a single object/factor combination) were converted to z-scores; the per-subject z-scores were then averaged across all subjects. Figure 3.8 depicts the results obtained from this experiment: Figure 3.8(a) depicts perceived importance as a function of object size; Figure 3.8(b) depicts perceived importance as a function of object location; Figure 3.8(c) depicts perceived importance as a function of the amount of object blurring; and Figure 3.8(d) depicts perceived importance as a function of object contrast.

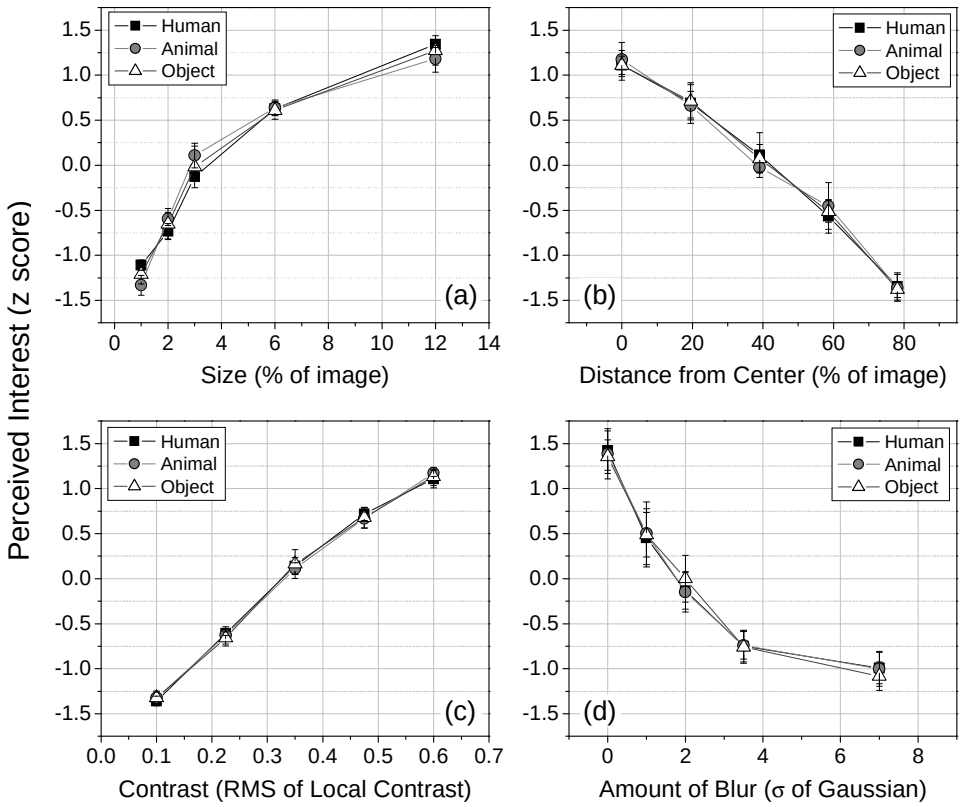

Figure 3.8*: Subjective ratings of visual interest. (a) Perceived interest vs. object size; (b) perceived interest vs. object location; (c) perceived interest vs. amount of object blurring; (d) perceived interest vs. object RMS contrast.*

The results of this experiment revealed that:

(1) As object size increases, perceived interest increases but exhibits diminished gains for larger sizes;

(2) As an object moves from the center of the image toward the image's edge, perceived interest decreases in a nearly linear fashion with distance;

(3) Blurring an object initially imposes a substantial decreases in perceived interest, but this drop in interest is relatively lessened for highly blurred images;

(4) As an object's RMS contrast is increased, perceived interest increases in a nearly linear fashion; and

(5) The relationship between each of the four factors and perceived interest is similar for all three categories (human, animal, non-human/non-animal object).

Thus this chapter presents us with a way of converting each of the factors influencing the ROI and its importance into the perceived score.

CHAPTER 4

Experiment-3

Now that the importance of each of the factors has been established we have tried to use them and formulate a semi automated algorithm for finding the importance maps of a gray scale image.

For this an experiment was conducted to find out the importance of different regions in an image based on viewer's ratings. Thus ideal importance maps were obtained for different images. Then each of these factors was optimized to give the best solution for predicting the importance maps.

4.1 Methods

4.1.1. Apparatus

Stimuli were displayed on a high-resolution, ViewSonic VA912B 19-inch monitor. The display yielded minimum and maximum luminance of respectively, 2.7 and 207 cd/m2, and an overall gamma of 2.9; luminance measurements were made by using a Minolta CS-100A photometer (Minolta Corporation, Tokyo, Japan). Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 46 cm through natural pupils under D65 lighting.

4.1.2. Stimuli

Stimuli used in this experiment were 15 gray scale images of different sizes. The original images were obtained from the calphotos database [33] and the stoch.xchng expert database [34]. These different color images were then converted to grayscale using the gimp software conversion for RGB to Gray. They were then resized such that the lower dimension was 512. Each of these images was then segmented into different regions. Figure 4.1 shows the different images used in the experiment and figure 4.2 shows the masks with the segmented regions for different images. The masks were hand drawn by the author.

Figure 4.1: *Fifteen different images used in the experiment*

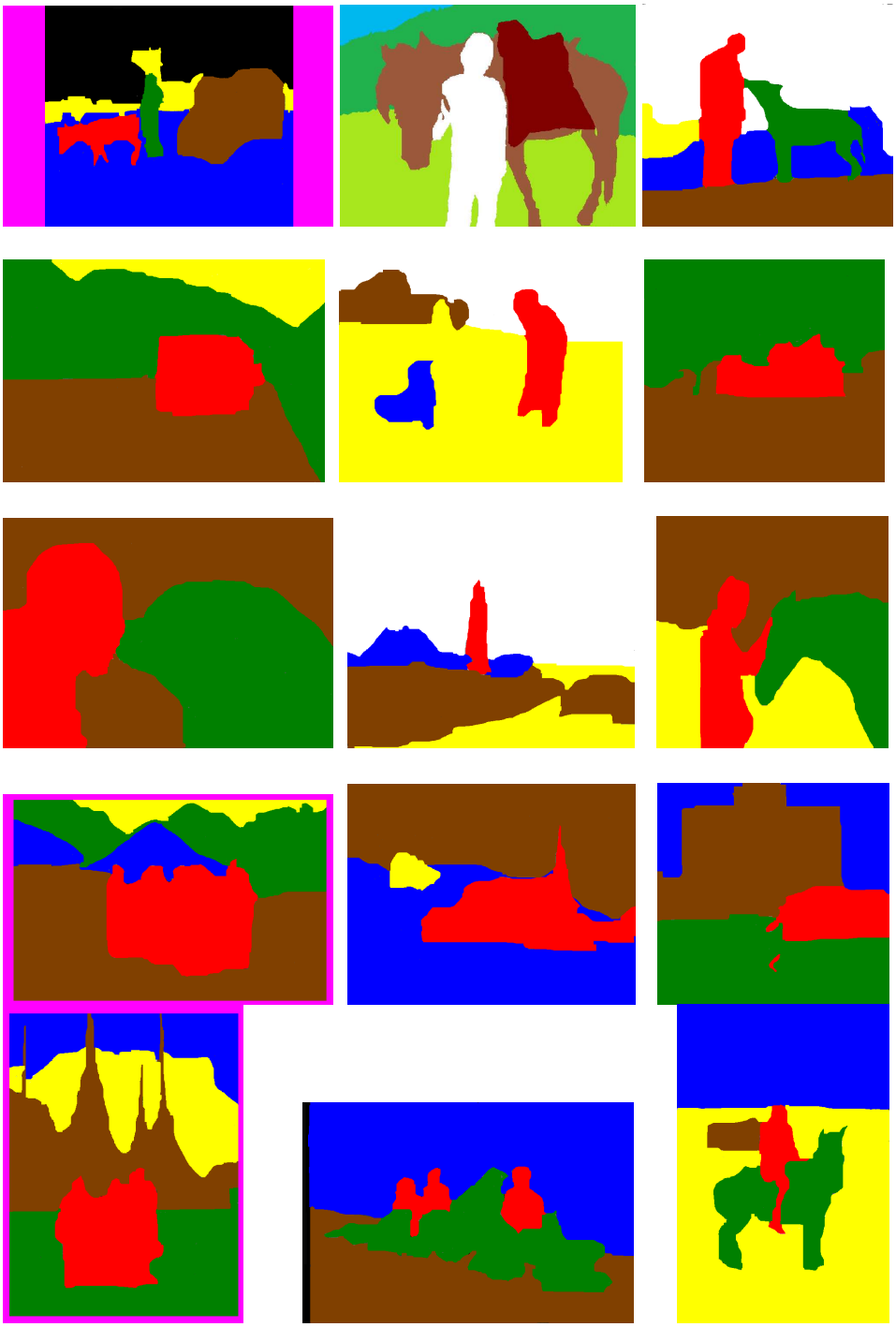

Figure 4.2: *Masks showing segmented regions for the images*

4.1.3. Procedure

Each experimental session began with three minutes each of dark adaptation and adaptation to a blank 19.0 cd/m2 display. Each of the original images was displayed on the screen and printout of the segments regions was handed out to the subject. The subjects were then given time to familiarize with the different regions in the image. Once they had the knowledge of different regions they were asked to rate the importance of each and every region in the image on a scale of 0-100 where 100 was the most important region. Ratings were reported verbally and recorded by a proctor. The time-course of experiment was not limited, however the majority of observers completed in approximately 50 minutes. We acknowledge that adaptation had likely occurred during the course of experiment, and that learning may likely have likely occurred.

4.1.4. Subjects

A total of 6 different subjects were used for this experiment. Two of the subjects were familiar with the purpose of the experiment. . Subjects ranged in age from 21 to 28 years. All subjects had either normal or corrected-to-normal visual acuity.

4.2 Results and algorithm

The raw scores for each subject on each set image were converted to z-scores; the per-subject z-scores were then averaged across all subjects. They were then normalized on a scale of 0 to 1. Thus every image had an ideal importance map with the importance of every region being proportional to the normalized value.

4.2.1. Algorithm

All the different regions in an image were divided into four categories. They were Humans, Animals, Objects and Background. The background included sky, ground and water. As observed from the previous experiment the different factors that were found to influence the ROI were selected. A new factor namely the Category is also introduced. Thus the five different factors to be used in this optimization process are

1) Size
2) Location
3) Contrast
4) Blur and
5) Category

The 15 different images used in the previous experiment were taken and for each of the images the following calculations were applied

1) The size of each and every region in an image was calculated using matlab code. This size was then converted as a percent of the total image.
2) The center of mass of each of the regions was calculated using the matlab code and that was assumed to be the location of that region
3) The local object contrast of the different regions in the image was calculated.
4) As far as blur was concerned the different regions were given the ratings for blur based on perceived blur by the student and the advisor.
5) The different regions were categorized into four different categories as explained above.

Thus the raw scores for each of these factors were obtained.

Then

1) Figure 3.8 (a) was used to convert the raw score for size into perceived interest.
2) Figure 3.8 (b) was used to convert the raw score for location into perceived interest.
3) Figure 3.8 (c) was used to convert the raw score for contrast into perceived interest.
4) Figure 3.8 (d) was used to convert the raw score for blur into perceived interest.
5) For finding the importance of each of the category the following procedure was applied. From the set of 15 images used in Experiment-3 all the regions containing humans were selected and the normalized value or the importance of each of these regions was averaged over all the humans. The value obtained was the perceived interest for that category. The same process was repeated for Animals, Objects and Background. The values obtained were as follows.

Humans	Animals	Objects	Background
1.000	0.896	0.584	0.047

Table 4.1: Importance of each of the categories.

Based on Figures 3.8(a), 3.8(b), 3.8(c), 3.8(d) equations were developed for each of these factors. These equations were the best possible fit that could be obtained.

1) **Size:** The best possible fit for the size was the sigmoidal fit. It was a Boltzman log data fit function. The equation was

$$y = \frac{A_1 - A_2}{1 + e^{(x - x_0)/dx}} + A_2 \qquad \textbf{3}$$

Where each of the factors is shown below

A_1	A_2	x_0	dx
-5.345	**1**	**-4.07596**	**3.02675**

Table 4.2: Parameters in equation three

The graph obtained is shown below

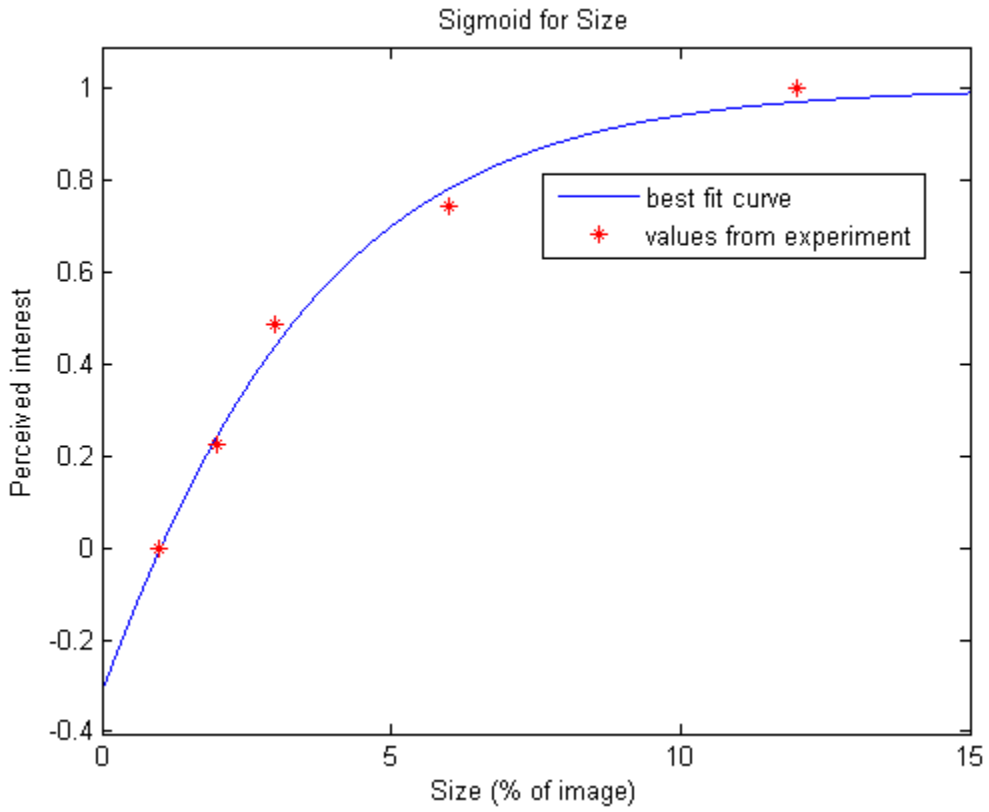

Figure 4.3: Sigmoidal fit for Size factor

The perceived interest for size was then calculated using a max operator.

Perceived Interest (Size) = max (0, y) [y is from equation 3]

2) Blur: The best possible fit for the blur was the sigmoidal fit. It was a Boltzman log data fit function. The equation was

$$y = \frac{A_1 - A_2}{1 + e^{(x - x_0)/dx}} + A_2 \qquad \textbf{4}$$

Where each of the factors is shown below

A_1	A_2	x_0	dx
1	**0.028**	**1.551**	**0.705**

Table 4.3: Parameters in equation four

The graph obtained is shown below

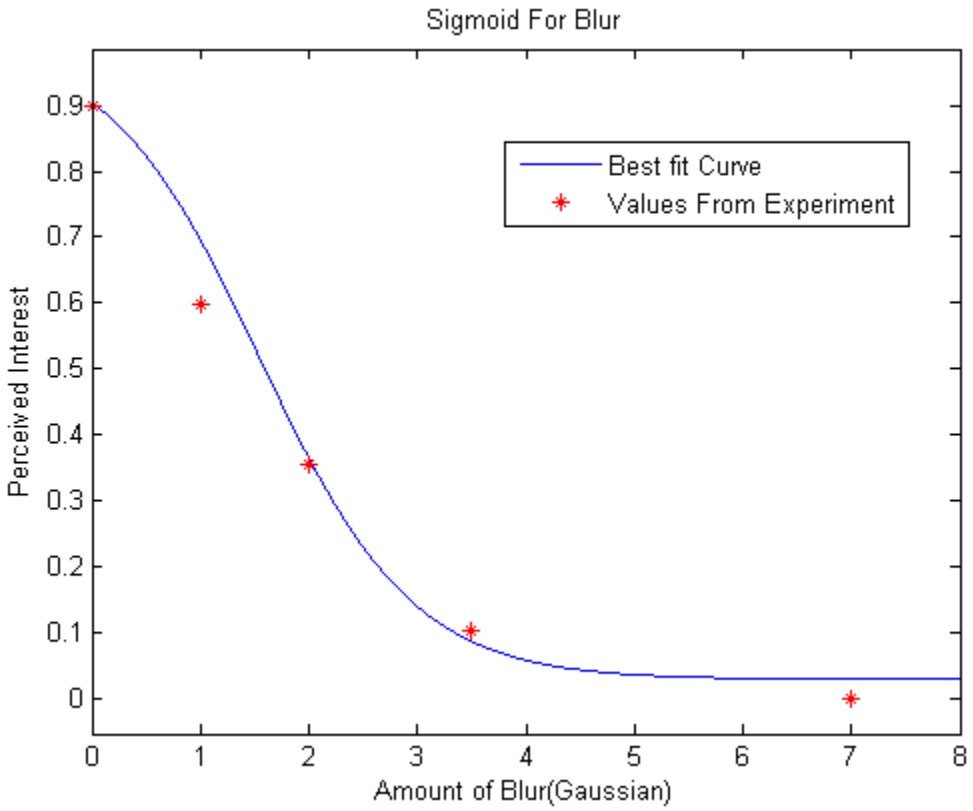

Figure 4.4: Sigmoidal fit for Blur factor

The perceived interest for blur was then calculated using equation four

Perceived Interest (Blur) = y [y from equation 4]

3) **Contrast:** The best possible fit for the size was the sigmoidal fit. It was a Boltzman log data fit function. The equation was

$$y = \frac{A_1 - A_2}{1 + e^{(x - x_0)/dx}} + A_2 \qquad \textbf{5}$$

Where each of the factors is shown below

A_1	A_2	x_0	dx
-0.185	**1**	**-0.275**	**0.106**

Table 4.4: Parameters in equation five

The graph obtained is shown below

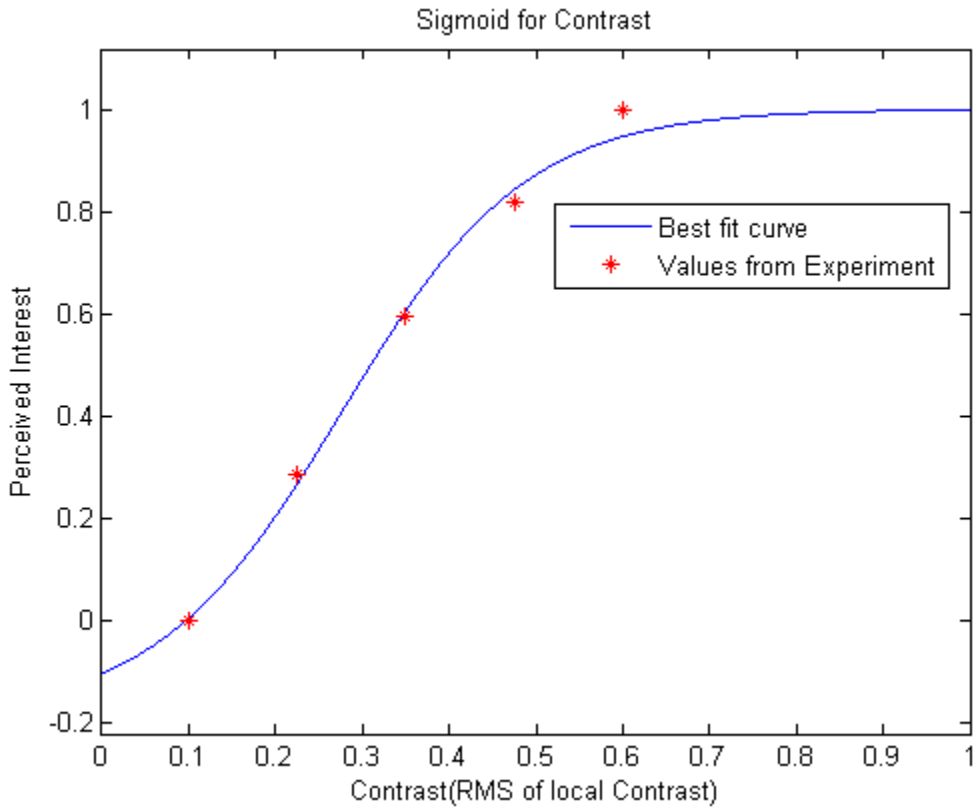

Figure 4.5: Sigmoidal fit for Contrast factor

The perceived interest for Contrast was then calculated using a max operator.

Perceived Interest (Contrast) = max (0, y) [y is from equation 5]

Now that we have the perceived interest for each of the five factors we have tried to develop an algorithm that would predict the importance map based on these factors of perceived interest. The basic goal was to produce an algorithm that would eventually predict the importance map for any given segmented gray scale image.

An optimization loop was run on all of these factors using a “Nelder-Mead Downhill Simplex Search”. This search was basically run to optimize for the weights of each of these factors. The search yielded the following results. The following weights were found to be optimal for each of the categories.

Size	Location	Contrast	Blur	Category
0.0321	0.1653	0.0976	0.2511	0.2527

Table 4.5: Optimal weights for the categories.

The Correlation R was reported as R = 0.9704, the graph is as shown below

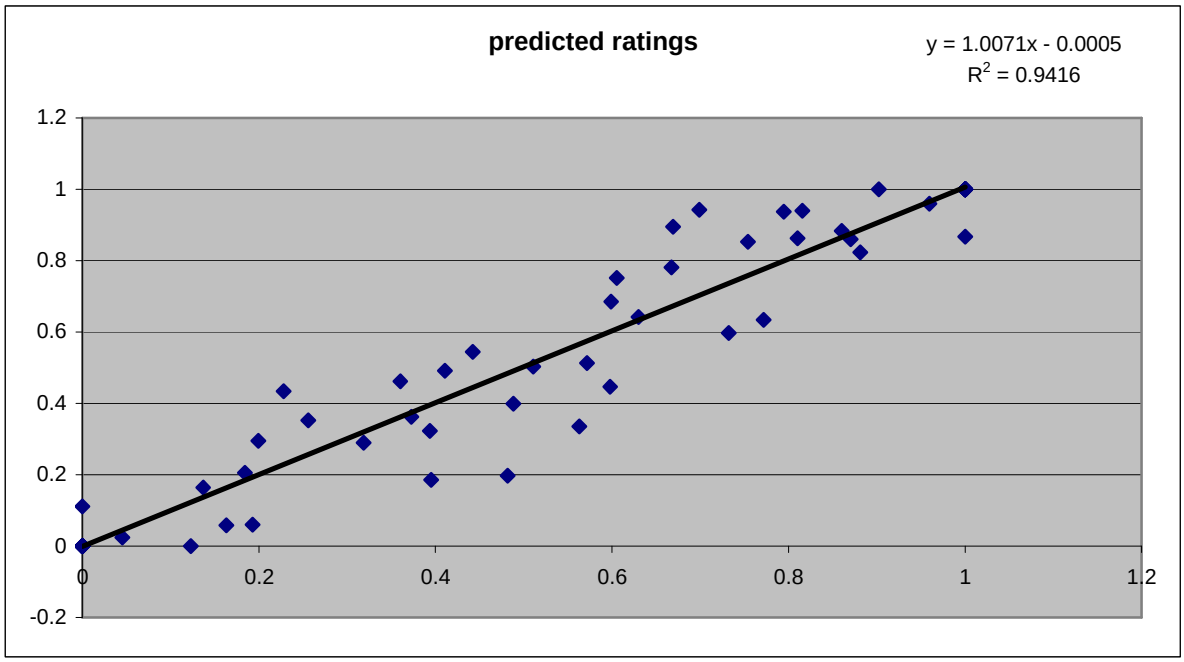

Figure 4.6: The graph showing the trend line and the equation for the best fit (proposed algorithm)

Thus the final equation for the importance of a region is given by

Importance = (*0.0321Size*) + (*0.1653**Location*) + (*0.0976**Contrast*) + (*0.2511**Blur*) + (*0.2527**Category*)** **6**

The same search was performed on Osberger's five factors and the results are as shown below

Contrast	Size	Location	Background	Shape
0.0845	0.0630	0.6420	0.2079	0.0326

Table 4.6: Optimal weights for Osberger's factors.

The correlation R was reported as R = 0.8377, the graph is as shown below

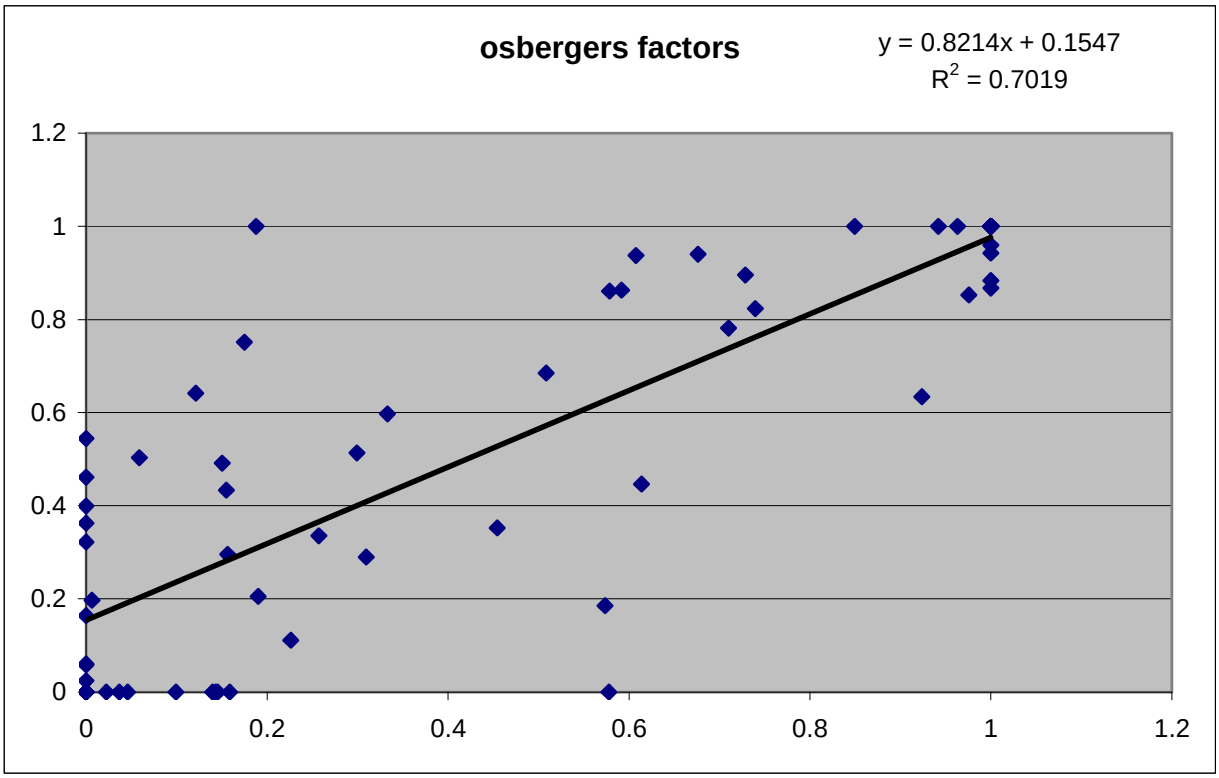

Figure 4.7: The graph showing the trend line and the equation for Osberger's algorithm

When the same search was performed without the category that is with just the four other factors the following results were observed

Size	Location	Contrast	Blur
0.0000	0.3377	0.1359	0.3682

Table 4.7: Optimal weights with no category.

The correlation R is R = 0.8499, the graph is as shown below

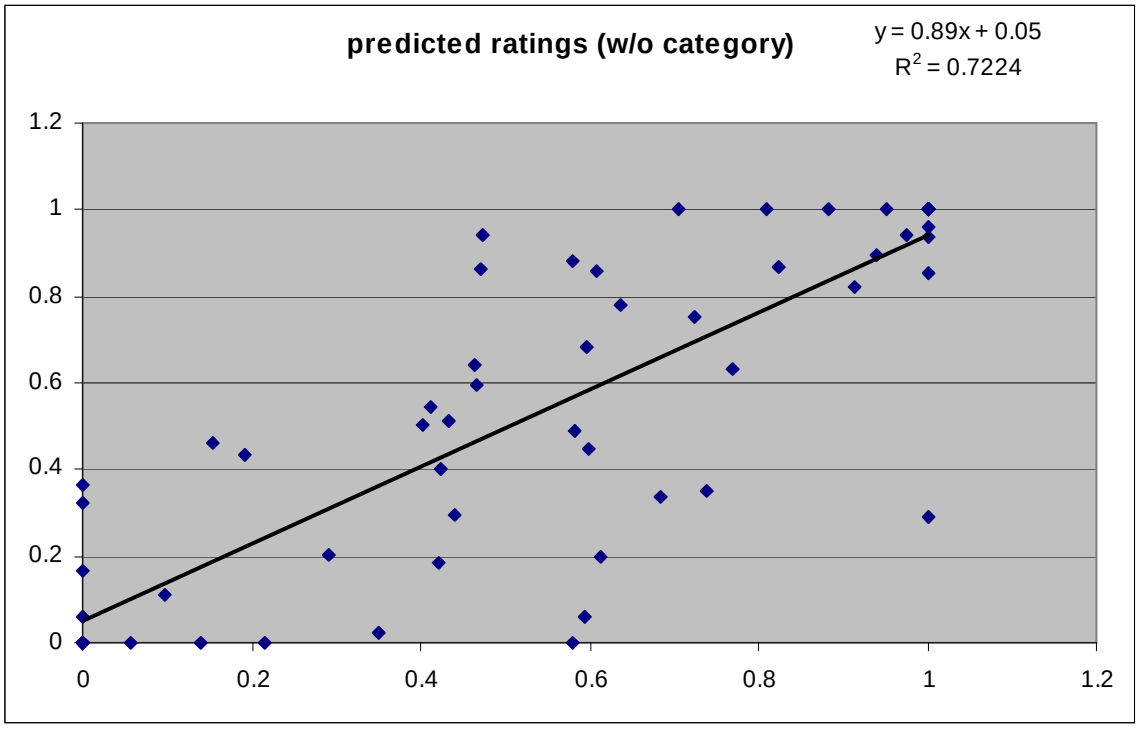

Figure 4.8: The graph showing the trend line and the equation for the algorithm with no category.

Thus it is clear from the above figures and correlations that the proposed algorithm is a lot better at predicting human ratings when compared to Osberger's algorithm. The correlation R for the proposed algorithm was 0.9704 which is a lot higher than the correlation for Osberger's algorithm which was 0.8377.

4.3 Comparison

Then finally a comparison has been between three algorithms

1) the algorithm suggested in the thesis
2) Laurent itti's algorithm based on biological model
3) Wilfried Osberger's algorithm based on his factors.

The results are shown below: when the given image is figure 1.4 (a)

Figure 4.9: Output of the proposed algorithm. The intensity of the regions being proportional to its importance, where white denotes the most important region. The importance of each of the regions is calculated from equation 6

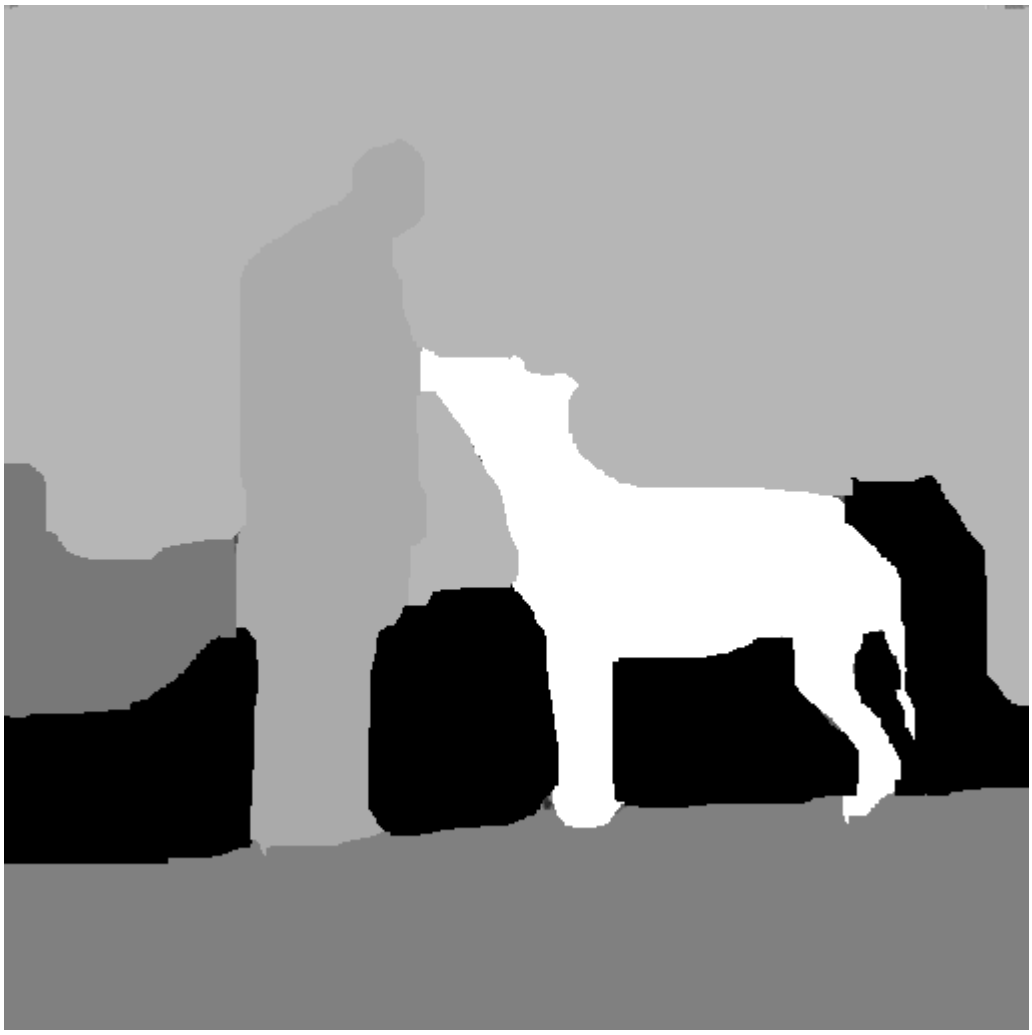

Figure 4.10: Output of the Osberger's algorithm (the different factors were applied to human segmented regions). The intensity of the regions being proportional to its importance, where white denotes the most important region.

Figure 4.11: Output of Laurent Itti's algorithm (no human segmentation). The intensity of the regions being proportional to its importance, where white denotes the most important region.

It is clearly evident from the above results that the importance maps given by the proposed algorithm are a lot better when compared to importance maps from other algorithms. The proposed algorithm is very good at predicting the human ratings.

CHAPTER 5

Conclusions and Future Work

The Importance of defining ROI for still images was clearly established by providing concrete evidence on the basis of experimental results. These results also suggested a way of proportionating the distortion between ROI and non-ROI. From these facts it has been established that knowledge of ROI has an effect on the image quality. The various factors influencing the ROI have been studied and the relationship between these factors and the perceived interest has been established.

Based on the results of the previous experiments a semi-automated algorithm has been proposed to produce an importance map for a segmented gray scale image. The results of the algorithm have been found to be satisfactory. A comparison was made between the proposed algorithm and two other widely used algorithm and the proposed algorithm has significantly better results.

The main drawback of this algorithm is the fact that it uses hand segmented regions and one of the factors “Category” is also user defined. Our future work is aimed at making both of these automatic. Thus a segmentation algorithm that is capable of segmenting the image into regions and identifying the category of different regions is being worked upon. This concept can also be effectively applied to color images and video.

References

[1] A.P. Bradley, Can region of interest coding improve overall perceived image quality? in: Proceedings of the Workshop on Digital Image Computing, Brisbane, Australia, 2003, pp. 41–44.

[2] A. P. Bradley and F. W. M. Stentiford, JPEG 2000 and Region of Interest Coding, *Digital Image Computing Techniques and Applications* (DICTA), Melbourne, Australia, January 2002

[3] A. P Bradley and F. W. M. Stentiford, Visual Attention for Region of Interest Coding In JPEG 2000, Submitted to *Journal of Visual Communication and Image Representation*, January 2002.

[4] Lijie Liu and Guoliang Fan, A New JPEG2000 Region-of-Interest Image Coding Method: Partial Significant Bit planes Shift, IEEE Signal Processing Letters 10 (2) (2003) 35–38.

[5]C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intel* **22**(9), pp. 970 982, 2000.

[6] Anthony Nguyen, Vinod Chandran, and Sridha Sridharan, "Gaze tracking for region of interest coding in JPEG 2000," Signal Processing: Image Communication 21 360 (2006) 359–377

[7] Eli Peli, "Feature Detection Algorithm Based on a Visual System Model**",** Proceedings Of The IEEE, Vol. 90, No. 1, January 2002.

[8] X. Marichal, T. Delmot, V. De Vleeschouwer, and B. Macq. Automatic detection of interest areas of an image or a sequence of images. In *ICIP*, pages 371–374, Lausanne,

Switzerland, Sep 1996.

[9] W. Osberger, N. Bergmann, and A. Maeder, “An automatic image quality assessment technique incorporating higher level perceptual factors,” Proc. IEEE Int. Conf. on Image Processing 3, 414–418 (1998).

[10]L. Itti, C. Koch, and E. Neibur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Mach. Intel* **20**(11), pp. 1254–1259, 1998.

[11]. J. H. van Hateren and A. van der Schaaf, “Independent component filters of natural images compared with simple cells in primary visual cortex,” *Proc.R.Soc.Lond. B* **265**, pp. 359–366, 1998.

[12] F. Kozamernik, V. Steinmann, P. Sunna, and E. Wyckens, “SAMVIQ: A new EBU methodology for video quality evaluations in multimedia,” *SMPTE Motion Imaging Journal* **114**, pp. 152–160, 2005.

[13] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, “Image and video quality assessment research at LIVE.” Online. http://live.ece.utexas.edu/research/quality/.

[14] E. Niebur and C. Koch. Computational architectures for attention. In R. Parasuraman, editor, *The Attentive Brain*. MIT Press, 1997.

[15] Taubman, D.S., Marcellin, M.W., 2001. JPEG 2000: Image Compression Fundamentals, Standards, and Practice. Kluwer Academic Press, Dordrecht.

[16]C. M. Privitera, M. Azzariti, and L. W. Stark, “Detecting regions-of-interest: the Mars Rover application,” in *Proc. Int.Workshop on Computer Vision, Pattern Recognition and Image Processing—JCIS’98*, Oct. 1998, pp. 215–218.

[17] C. M. Privitera, L. W. Krishnan, and C. M. Stark, “Clustering algorithms to obtain regions of interest: a comparative study,” *Proc. SPIE*, pp. 634–643, 2000.

[18] C. M. Privitera and L. W. Stark, “Evaluating image processing algorithms that predict regions of interest,” *Pattern Recognit. Lett.*, vol. 19, no. 1, pp. 1037–1043, 1998.

[19] R. Grosbois, D. Santa-Cruz, T. Ebrahimi, New approach to JPEG 2000 compliant region of interest coding, in: Proceedings of the Applications of Digital Image Processing XXIV, vol. 4472, San Diego, USA, 2001, pp. 267–275.

[20]A. Nguyen, Importance prioritised image coding in JPEG 2000, Ph.D. Thesis, School of Engineering Systems, Queensland University of Technology, Brisbane, Australia, January 2005.

[21] A. Nguyen, V. Chandran, S. Sridharan, Importance prioritization in JPEG 2000 for improved interpretability, Signal Process.: Image Commun. 19 (10) (2004) 1005–1028.

[22] A. Nguyen, V. Chandran, S. Sridharan, Visual attention based ROI maps from gaze tracking data, in: Proceedings of the International Conference on Image Processing, Singapore, 2004, pp. 3495–3498.

[23] A. Nguyen, V. Chandran, S. Sridharan, GazeJ2K: A gaze influenced JPEG 2000 image coder, in: Proceedings of the Workshop on the Internet, Telecommunications and Signal Processing, Adelaide, Australia, 2004, pp. 18–23.

[24] A. Nguyen, V. Chandran, S. Sridharan, Gaze-J2K: Gaze influenced image coding using eye trackers and JPEG 2000, J. Telecommun. Inform. Technol. 1 (2006), in press.

[25] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, Importance coding of still imagery based on importance maps of visually interpretable regions, in: Proceedings of the International Conference on Image Processing, vol. 3, Thessaloniki, Greece, 2001, pp. 776–779.

[26] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, Importance assignment to regions in surveillance imagery to aid visual examination and interpretation of compressed images, in: Proceedings of the International Symposium Intelligent Multimedia, Video and Speech Processing, Hong Kong, 2001, pp. 385–388.

[27] A. Nguyen, V. Chandran, S. Sridharan, R. Prandolini, JPEG2000 region of interest coding—a hybrid coefficient scaling and code-block distortion modulation method, in: Proceedings of the Australasian Workshop on Signal Processing and Applications, Brisbane, Australia, 2002, pp. 59–62.

[28] W. Osberger, Perceptual vision models for picture quality assessment and compression applications, Ph.D. Thesis, School of Electrical and Electronic Systems Engineering, Queensland University of Technology, Brisbane, Australia, March 1999.

[29] C. Privitera, L. Stark, Focused JPEG encoding based upon automatic pre-identified regions-of-interest, in: Proceedings of the Human Vision and Electronic Imaging IV, vol. 3644, San Jose, USA, 1999, pp. 552–558.

[30] The Mathworks Inc., Matlab User Guide: Statistics Toolbox, Version 6 (R14), 2006.

[31]Z. Wang, A.C. Bovik, Bitplane-by-bitplane shift (BbBShift)—a suggestion for JPEG2000 region of interest image coding, IEEE Signal Process. Lett. 9 (5) (2002) 160–162.

[32] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[33] http://calphotos.berkeley.edu/browse_imgs/people.html

[34] http://www.sxc.hu/category

[35] A. Olmos and F. A. A. Kingdom, “McGill Calibrated Colour Image Database,” Http://tabby.vision.mcgill.ca.

[36] B. Moulden, F. A. A. Kingdom, and L. F. Gatley, “The standard deviation of luminance as a metric for contrast in random-dot images,” Perception 19, 79–101 (1990).

[37] Hans van Hateren database; last updated: April-05, http://hlab.phys.rug.nl/archive.html

[38] Microsoft Clip Art, http://office.microsoft.com/en-us/clipart/default.aspx

VITA

Kadiyala Vamsi Krishna

Candidate for the Degree of
Master of Science

Thesis: IMPORTANCE OF DEFINING ROI: A SEMI-AUTOMATED ALGORITHM FOR PREDICTING IMPORTANCE MAPS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in India, on May 23, 1984.

Education: Received the B.Tech degree from Andhra University Visakhapatnam, India, in 2005, in Electronics and Communicatons Engineering; Completed the requirements for the Master of Science degree with a major in Electrical Engineering at Oklahoma State University in Oct 2007.

Experience: Research Assistant at Oklahoma State University from August 2006 to September 2007.

Professional Memberships: IEEE Student Member.

Name: Kadiyala Vamsi Krishna Date of Degree: December, 2007

Institution: Oklahoma state University Location: Stillwater, Oklahoma

Title of Study: **IMPORTANCE OF DEFINING ROI: A SEMI-AUTOMATED ALGORITHM FOR PREDICTING IMPORTANCE MAPS**

Pages in Study: 67 Candidate for the Degree of Master of Science

The importance of defining ROI in gray scale images is clearly established. A new way of proportionating the distortion between ROI and non-ROI is found out. The various factors influencing the ROI have been studied and the relationship between these factors and the perceived interest has been established. A new Algorithm has been proposed to produce an importance map for a given gray scale image. The algorithm takes into account various factors like Size, Location, Blur and Contrast. The results of the algorithm have been found to be satisfactory. A comparison was made between the proposed algorithm and two other widely used algorithm and the proposed algorithm has significantly better results.

ADVISOR'S APPROVAL: Dr Damon Chandler